William L J Xu-Doeve

Introduction to the

Measurement of Internal and International Migration

ANRC Publishing

Introduction to the

Measurement of Internal and International Migration

William L J Xu-Doeve

Published 2008 by ANRC Publishing, the Netherlands

Library Cataloguing-in-Publication Data
Xu-Doeve, William L J, 1953-.
Introduction to the Measurement of Internal and International Migration.
/ W.L.J. Xu-Doeve
p. xxv + 138. 29.7 × 21.0 cm. tables, figures
Includes bibliographical references
ANRC Publishing, 2008
ISBN: 978-90-8802-002-5 (eBook edition)

Subject Area 1. Demography. 2. Population Studies. 3. Geography. 4. Planning. 5. Economics. 6. Sociology. 7. Statistics.
Keywords 1. Internal migration. 2. International migration. 3. Demography. 4. Methodology. 5. Demographic measurement. 6. Migration data specifications. 7. Migration data collection. 8. Migration data sources. 9. Population census. 10. Data processing.

NUR 755

Copyright © 2007, 2008 William L J Xu-Doeve. All rights reserved.

Without permission in writing from the copyright owner this work may not be translated, adapted, republished, redistributed, reproduced, printed, reprinted or copied in whole or in part in any form or by any electronic, magnetic, optical, mechanical or other means now known or hereafter invented or developed including but not limited to photocopying, recording or use in connection with any form of information storage, retrieval or distribution system. Please refer to the final page of the Author's Preface for contact information.

Permission is granted to the purchaser of the digital eBook edition of this work to print on paper one copy of the eBook purchased and to digitally store one copy of the eBook purchased in such a manner that the digital eBook purchased is and remains properly secured against any infringement on the copyright including against any violation of the terms, conditions, limitations and restrictions set out above.

Permission is granted to quote brief passages from this work in connection with reviews or scholarly analysis provided the customary acknowledgement of the source is given.

Every effort has been made to ensure that the information and advice in this work is true and accurate at the time of going to press. However, neither the publisher nor the author can accept any legal responsibility or liability for any errors or omissions.

EDITORIAL FOREWORD

Migration has become a key issue of concern for policy makers in many countries. For many years, *internal migration* and sustained high rates of urban growth, for example, have been posing major problems in terms of the provision of infrastructure, housing, sanitation, health care and education in many developing nations. More recently, the question of *international migration*, too, has risen to the top of the policy agenda in developed and developing countries alike.

Sound *evidence-based* policy making on these issues requires timely and reliable information. Yet, however important the migration question may be, from a scientific point of view the traditional *methods of measuring* actual migrant flows, migrant stocks, and their dynamics over time have long been characterized by important *methodological weaknesses*. And the *resulting information* on actual migration processes and on true migrant numbers is typically characterized by a very considerable and, many would argue, unacceptable degree of *empirical uncertainty*.

More recently, however, *new and powerful event-based methods of measuring migration* have emerged. These are methods which are methodologically sound and which, in addition, result in comprehensive and fully-detailed information on actual migration processes and migrant numbers.

For many years now, ANRC Consulting has been providing statistical capacity building services in developing and emerging economies [1]. Migration -- both the collection of data, and methods of measurement and analysis -- is one of several

1 For more details, see: http://www.anrc-consulting.com/en/csr.html

key areas of expertise. Some two years ago, the demands from this programme led to the publication of an in-depth manual dedicated to these new methods of measuring migration [2]. This is a manual which is aimed at more advanced users who already have a reasonably thorough grasp of the formal methods of demography and of the materials of population statistics.

Clearly, this left an unmet need for a textbook of a more introductory nature. The present book serves to fill this void. It has been designed in particular for users who are interesting in the actual measurement and analysis of migration, but who do not have a background in formal demographic analysis, or who have no interest in such formal rigour.

The material presented in this book is focused on enabling *direct practical application*. It is organized in two chapters.

> The first chapter provides a non-technical introductory overview of modern *methods of measuring migration*.
>
> The second chapter deals with the implications for *migration data specifications*, *data sources*, *data collection*, and *data processing*.

These two chapters have been designed to be reasonably *self-contained*, and they can be read independently. Each chapter is supported by its own abstract and its own detailed table of contents. In addition to the book page numbers in the top right-hand corner of each page, the chapters also have chapter page numbers in the bottom right-hand corner.

It is recommended that *new users* of this book first read the author's preface starting on page *vi*. This preface places the material presented in this book in context, and it provides important background information.

This new book is an ideal first reference for *national and international statistical offices* and for *research institutions* interested in migration data collection and in the measurement of internal and/or international migration.

In addition, it is particularly suitable to serve as an introductory textbook on methods of measuring migration in *undergraduate* and *post-graduate academic programmes* in subject areas such as demography, population studies, statistics, economics, sociology, geography, and urban and regional planning.

---

2 Xu-Doeve, W L J (2006) *Methods of Measuring Internal and International Migration*. ANRC Publishing. ISBN 978-90-8802-001-8

Finally, with its emphasis on practical application and its clear guidelines for operational implementation, this book is also ideally suited as a training manual on the state of the art in methods and practices for *professional training programmes* and other *dedicated short courses* specifically focusing on internal and/or international migration.

Many countries worldwide are currently planning and preparing for a new population census. This also makes the publication of this book timely, because it pays very considerable attention especially to the use of population census data for the measurement of migration:

One of the key objectives of the book is to provide sound *operational guidelines* in preparation for migration data collection and for the measurement and analysis of migration in the *2010 global round of population censuses*.

Louise Amstein
Editor

AUTHOR'S PREFACE

This book presents an introduction to events-based methods of measuring internal and international migration. These are new and powerful methods which, for the first time in the study of migration, are able to produce fully comprehensive and in-depth insights into ongoing and historical migration processes.

In this preface we shall very briefly discuss a number of topics which place the material presented in this book in a broader context, and we shall highlight some of the key issues which have guided the selection and presentation of this material.

Specifically, we begin by placing these new methods of measurement in perspective by contrasting them with familiar and widely used traditional methods of measuring migration. Next, and building on this, we make some cautionary remarks about commonly used methods of projecting and forecasting populations which incorporate population redistribution through migration.

In the third section, prerequisites, we then list the recommended background for users of this book. Next, section 4 of this preface discusses our approach to bibliographic referencing in this book.

In most countries, the periodic population census will be the only available source of comprehensive data on migration processes. However, the collection of adequate migration data in population censuses is a challenging task. Section 5 provides some background information concerning the recommendations in this book as regards how best to meet these challenges.

Finally, we conclude this preface with a historical note. This aims to place the approach taken, and the manner in which it is presented, in context.

1 The Methods in Perspective

Events-based methods of measurement are demographic methods that centre on *migratory events* or *migratory moves* as they occur in the *life histories of persons*.

Thus, there are two key distinguishing elements that characterize these methods: First, the primary focus is on the occurrence of *demographic events* (migratory moves), rather than on persons (migrants). Second, the methods are based on a rigorous *life history* or *cohort* framework of analysis, and they maintain this cohort perspective throughout.

To avoid any misunderstanding, we emphasize that information on migrants is not ignored or considered irrelevant. Quite to the contrary. However, moves are the starting point of the analysis, and full information on persons (migrants) is then *derived* as a direct result of this analysis.

These event-based methods of measurement differ fundamentally from common traditional methods of measuring migration. The most often encountered traditional methods of measuring internal and international migration broadly fall into one of three main categories:

The first and most common of the traditional methods of direct measurement is *person based*. Here, a person is defined as a migrant on the basis of the value of some personal attribute (characteristic). Three of the typical attributes that are frequently used include place or country of birth, place or country of residence one or five years earlier, and nationality or citizenship. If the attribute considered implies that a person has experienced at least one migratory move, then the person in question is regarded a migrant. The measurement of migration is then simply reduced to the *counting of these migrants*.

The destination of the migrant in question is interpreted as the current place of residence, and this raises no issues of principle. However, the origin of the migrant is simply taken as the origin implied by the observed value of the attribute. Following the previous three examples of attributes, the origin would then be considered to be, respectively, the place or country of birth, the place or country of residence one or five years earlier, or the country of nationality / citizenship.

Neither the true underlying *geographic trajectory*, nor the *number of actual moves* made, nor the *timing of any of these moves* is captured or explored. And, inevitably, *several categories of migrants* are *not actually considered migrants*. Among others, this includes all return migrants, and also all those international migrants who have adopted the nationality or citizenship of the destination country.

Clearly, since the actual underlying migration behaviour itself, that is, the actual migratory moves -- both by their timing and by their origin and destination -- remain unobserved, the insights that can be obtained from person-based measurement are necessarily limited and incomplete.

As already implied above, one important contributing cause of this, of course, is the fact that a person can experience multiple moves during any given time interval.

A second common traditional approach is *indirect estimation*. Indirect estimation is *also person based*, and it is applied when there are no adequate data on migration itself. Instead, *net numbers of migrants* over the time interval considered are *inferred* as the *unexplained residual* after accounting for regional population change over this time interval due to ageing and due to known or estimated patterns of mortality and fertility.

Clearly, even if there would be no estimation errors, indirect estimation results in information on actual migration processes that is even less complete and less informative than the direct count of numbers of persons defined as migrants on the basis of some attribute indicative of earlier migration behaviour.

A third and final, albeit somewhat less common, traditional approach to the measurement of migration is the computation of *empirical annual migration rates*, in a manner similar to the computation of the well-known empirical period mortality rates ${}_nM_x$. Contrary to the above person-based approaches, the measurement of such empirical annual occurrence / exposure rates is a method which does consider individual migratory events or moves.

Specifically, for the numerator of the rate, or the "occurrence", this approach requires a count of the number of migratory moves that have occurred during a given year. For the denominator of the rate, or the "exposure", one uses a direct count or, more usually, some estimate of the number of person-years that were exposed to the risk of experiencing a move during that year. Generally, the necessary data for the computation of such rates are readily available only in countries which maintain a continuous population registration system in which information on migratory moves is recorded. This explains why this approach to measurement is less commonly encountered than person-based methods.

However, as we shall see in some detail in chapter 1, traditional empirical annual occurrence / exposure rates of the type ${}_nM_x$, whether they be mortality rates or migration rates, are mathematically, methodologically, and often also empirically flawed, and limited.

In addition, the generalization of this type of rate from mortality to migration fails adequately to recognize the importance of the fact that, contrary to the event of death, the event of migration can occur more than once during the life time of a person. Traditionally computed period rates of the type ${}_nM_x$ do not capture and cannot properly express this essential cohort perspective.

This is not to say, though, that migration rates, as key indicators of the intensity of migration processes, are not important. Quite to the contrary, in fact: As we shall see in chapter 1, they turn out the be a pivotal concept that is central to the measurement of migration. However, instead of traditional period occurrence / exposure rates of the type ${}_nM_x$, in this book we shall encounter *time-continuous instantaneous cohort migration rates*, commonly denoted by $\mu(t)$.

These latter instantaneous cohort rates do not suffer from any of the conceptual, mathematical, methodological and empirical deficiencies that are invariably associated with traditional period rates. In addition, as we shall see, they can be simply and reliably derived, and not only from continuous population registration system data, but equally from standard population census data and from random sample survey data.

2 Projections and Forecasting

Note: This section 2 of the preface requires a basic familiarity with the principles and methods of population projections and population forecasting. Readers who have no background or interest in this area can skip this second section and proceed to section 3 without any loss of continuity.

This book focuses on *methods of measuring* and *analysing* migration. In other words, it centres on the development of *empirical information*. It *does not deal with* methods of and models for population *projections* and population *forecasting*.

However, the projection and forecasting of populations that are disaggregated by place of residence, such as urban and rural, regional, and local projections and forecasts, is a matter of considerable importance for evidence-based policy making and planning. Methodologically, it is a topic that belongs to a well-developed and related but separate subject area, nowadays commonly called multistate demography [1].

Clearly, having an adequate *benchmark data set* on the population structure and the components of population change, including, of course, migration, is an elementary prerequisite for such spatially disaggregated population projections and forecasts. Although this topic is beyond the scope of the present book, a *note of caution* is appropriate, here.

Modern multistate population projection methods were first developed in the 1970s and 1980s as mathematical generalizations of traditional or classical cohort component methods that were formulated using Leslie matrices. A Leslie matrix encapsulates the components of change which determine how a population is affected by demographic processes as time and age progress. The classical Leslie matrix incorporates only mortality and fertility, while the multistate Leslie matrix also incorporates population redistribution through internal migration. We note here that this matrix formulation is merely a convenient and insightful notational tool; it is not essential, and what follows applies equally to other more traditional ways of formulating cohort component projection methods.

As commonly used as a population projection tool in applied demography, the name "cohort component method" is, however, something of a misnomer. Usually in practice, the projections are actually of a *hybrid* type, incorporating a mixture of *both* period or cross-sectional analysis *and* cohort or longitudinal analysis.

It is true that the projection process itself, in linear algebra conveniently represented by the multiplication of a Leslie matrix and a population vector, does project populations properly along cohort lines.

However, in parallel with the application of the classical Leslie matrix, in applied practice the multistate Leslie matrix *itself* is commonly *operationalized* based on period life table analysis or on similar cross-sectional principles for the survival and geographical redistribution elements. The fertility elements of the matrix, too, normally derive from period fertility data.

Period or cross-sectional life table and fertility schedule analysis both treat the experience of a range of different cohorts of successive age, each empirically observed over the same short interval of time, as if it were the life-time experience

---

1 Rogers (1975), listed in the references to chapter 1, is an early attempt at a comprehensive account.

of one single cohort. Such a resulting single cohort is, of course, merely a notional or imaginary construct. It is often called a *synthetic cohort*, so as to distinguish it clearly from any real empirical cohorts.

Using such synthetic cohorts is an approach which has become firmly embedded in demographic training programmes, as well as in the mind-set and applied routine of demographers.

Clearly, as a theoretical construct, there is nothing intrinsically wrong with the analytical device of a synthetic cohort. Quite the opposite, in fact. Synthetic cohorts have, for example, proven their value in the theory of stationary and stable populations, yielding powerful theoretical insights regarding the dynamics of populations that are subject to time-invariant (that is, constant or fixed) schedules of demographic change.

However, it would be wrong a priori to assume that the experience of such a synthetic cohort is, or may simply be taken as, the actual empirical experience of some real cohort. Yet, effectively this is what is commonly being done in applied cohort-component-based population projections and forecasting.

Empirically, this approach is indefensible, and most particularly so when migration is incorporated as one of the components of population change.

Cross-sectional or period analysis does not adequately recognize the fundamental importance of a longitudinal or cohort perspective in the analysis of migration. As a life-cycle-related phenomenon *and* against the background of a dynamic socio-economic, political, cultural and physical environment, the experience of migratory events, that is, the *making of migratory moves*, is a function *both* of the <u>*age* of the person</u> *and* of the <u>*time* in which this person lives</u>.

The relationship between *age* and the propensity to make a migratory move can be expressed by a cohort's observed age-specific migration schedule or pattern, just as, for instance, its propensity to die can be expressed by its observed age-specific mortality schedule.

There tend to be regularities in such migration schedules which tie in with *life-cycle events*. It is common, for instance, to see a local maximum or peak at the start of young adult life, which is related to job seeking, entry into higher education, military service, and so on. A peak at the youngest ages can be common as well. Normally this then is a derived local maximum, explained by young children migrating with their young adult parents. A less pronounced retirement peak can also sometimes be seen.

However, such regularities need not be general. Among other things they are, for example, dependent on the migration-defining areas studied. And, importantly, they can be modified by prevailing socio-economic, political and other *conditions* *at the time*.

Importantly, this relationship between prevailing conditions at the time and the propensity to make a migratory move causes such observed migration schedules to vary from cohort to cohort. In other words, cohort migration schedules are *not stable* from one cohort to the next.

It is useful to put this in perspective: Here we have a fundamental difference with, for instance, mortality. Mortality schedules are to a significant extent governed by stable biological factors, and the schedules tend to change only fairly gradually from cohort to cohort due to factors such as improvements in sanitation and advances in health care as time progresses. Specific events or conditions occurring at some given time only, such as wars, famines and epidemics, tend to be incidental factors underlying schedule change, rather than the structural factors.

Migration schedules, on the other hand, lack any similar stable determinants. Here, dynamic circumstances and conditions prevailing from time to time are structural factors in changing schedules. And successive cohorts each experience these same circumstances and conditions at a different phase during their respective life cycles.

Scientifically, therefore, there can be no a priori justification of the assumption that the observed past migration experience of older cohorts will in future be similarly replicated by younger cohorts. Nor can there be such justification of the assumption that older cohorts have in the past behaved in a manner similar to the observed current behaviour of younger cohorts.

The lack of stability in migration schedules from cohort to cohort has two key implications in the population forecasting process.

First, the establishment of a *benchmark database* which is adequate as a point of departure and which gives sufficient guidance for the formulation of well-founded forecasting scenarios, places significant demands on empirical migration data collection and on the measurement, analysis and interpretation of relevant ongoing and historical migration processes. Put more specifically, the forecasting of populations where migration plays a role as one of the components of population change, requires the judicious empirical study of past migration behaviour along cohort lines, that is, as a function not only of age but also of time.

And, second, the design and formulation of the *forecasting tools* that will be used must be able explicitly to recognize and represent well-founded scenarios of expected specific longitudinal behaviour of the individual cohorts making up the population as time and age progress. The elegant simplicity of the cohort component model in itself cannot serve as an excuse to legitimize its use as an applied forecasting tool or to validate its output.

We note that before the development of multistate demography, methodologically less sophisticated methods and models were used for sub-national population projections and forecasting, methods which are still sometimes used today. In addition, applied models may have been modified so as not to be closed to external migration.

This note of caution, emphasizing the importance of proper prior longitudinal cohort data analysis and of explicitly recognizing that migration schedules are dependent both on age and on the time in which populations live, applies equally to the use of all such other projection and forecasting methods and models.

Finally, we stress that, contrary to the cross-sectional principles and/or practices commonly underlying the application of multistate and earlier population projection methods and models, the *methods of measurement presented in the present book* are *fundamentally and rigorously based on proper cohort analysis*.

Therefore, they provide a methodologically sound basis for the establishment of an adequate benchmark data set that can serve as a maximally-informed and dependable point of departure for the development of empirically well-founded scenarios of the dynamics of the future migration behaviour of the cohorts which make up the population under study.

3 PREREQUISITES

The measurement of internal and international migration processes, the subject of chapter 1, and the collection of appropriate migration data, the theme of chapter 2, are both issues which require careful thought and adequate prior consideration. The material presented in this book has been written so as to serve as an *introduction* to these subjects.

At the same time, a primary aim has been to provide at least enough information to facilitate actual practical application and implementation of the methods discussed. Therefore, we have endeavoured to ensure throughout that the accessibility and clarity of this introductory textbook has not been achieved by simply sacrificing either the necessary *precision* or sufficient attention to *detail*.

As regards *prerequisites*, it is recommended, though not necessary, for users of this book to have a basic familiarity with the standard demographic paradigm, and in particular with the concept of a life table. The final section of chapter 2 lists one of several available textbooks which the reader may find useful as a source of background information in this respect.

Further, we have chosen not to treat the actual development itself of the underlying mathematical theory in this book. In order to help building bridges to more advanced work, only some key resulting formulae are given in chapter 1 for interested readers. However, if so desired, these can simply be taken at face value without any loss of coherence or of continuity. As a result, no special mathematical skills are required for this book.

Operationalization of the methods of measurement does require one aspect of elementary statistics, namely, the topic to which statisticians commonly refer as "curve fitting". However, the level necessary for this book has deliberately been kept limited. And consequently, the material presented is readily accessible to users who have taken no more than a standard introductory course in statistics.

4 Bibliographic References

The material presented in this book has been very extensively peer reviewed. There are, however, two issues where some reviewers have made suggestions which, after due consideration, we have opted not to adopt. The first relates to bibliographic references, and we shall take up this topic in the present section. The second issue concerns our discussion of population censuses in chapter 2. We shall motivate our position as regards that matter in the fifth section of this preface.

For this introductory textbook we have chosen to limit the bibliographic referencing of related materials to three carefully selected types of publication.

First, there are the publications which we unreservedly recommend as useful sources of supplementary and/or contrasting material for all users of this book. These publications are easily identified, because they are listed and discussed in the final section of chapter 2.

Then there are a number of standard textbooks to which reference is made in chapter 1 in the context of the discussion on specific topics. They are included in the references to that chapter. The user may or may not find it interesting to consult these for further background information on the particular topics in question; the text of chapter 1 where we refer to these books provides useful guidance.

Finally, also in the references to chapter 1, there are a number of publications which are not generally well known. These have been included because, considered from a scientific point of view, they are of significant historical interest in the context of the present book. Most users, however, will likely be quite satisfied to take note of the existence of these publications without actually consulting them.

Some academic reviewers have suggested significantly to extend the number of references to include further literature materials. However, given the introductory nature of this book, it is questionable whether this truly serves a useful purpose.

Historically, the methods presented in this book were developed starting with a clean sheet of paper. Section 6 of this preface gives some more details on their design and development. Further, for several years in the early 1980s these methods were successfully taught to undergraduate students at the University of Manchester with no more than a few teaching hours of prior exposure to demographic methods and with no specific background related to the issue of migration.

Clearly, therefore, apart perhaps from a basic familiarity with demographic methods and materials, these new methods do not necessarily require any extensive knowledge of existing literature dealing with studies in internal or international migration.

The question then arises whether readers would actually benefit from more extensive referencing. Some undoubtedly would, such as, for instance, academic researchers, as well as post-graduate and doctoral students. However, more than likely these readers will have the necessary skills to collect such additional materials themselves.

For other users, such as, for example, for officers in national statistical offices or for undergraduate students, especially in the developing world, a much more comprehensive collection of references might instead well prove to be less than encouraging, or even frustrating because of their likely unavailability locally.

For this book we have chosen to side with the latter group among our target readership. However, this does, of course, not mean that we abandon other readers: They are referred to our more advanced companion book on methods of measuring migration[2] which does contain ample references to related methodological materials and to interesting empirical work.

---

2 See footnote 2 in the Editorial Foreword.

5 The Future of Population Censuses

A second issue raised by some referees (see section 4 of this preface) centred on our approach to population censuses, and in particular on our recommendation to *extend the recording of migratory moves* in population censuses *beyond merely the single most recent migratory move*. This is, however, a key recommendation. As we shall explain in the present section, limiting migration data collection to so-called last move data leaves immense gaps in our knowledge and understanding of actual true migration processes, gaps which would be considered totally unacceptable if it would concern not migration but fertility or mortality.

As will become clear in chapter 2, for most countries the periodic population census is the only suitable source of data that can yield comprehensive information on the ongoing processes of internal and international migration that contribute to shaping the nation. The most common way to obtain information on migratory events (migratory moves) in a population census is by *retrospective questioning* of population members on their duration of residence in their current place / country of residence and on the associated place / country of previous residence, if any migratory moves have indeed been made.

Traditionally, and still to this day according to the current international standard recommendations for population censuses, such questioning is limited to no more than the most recent migratory move only. If a person has made any earlier moves, then this remains unobserved.

Clearly, this limitation is a wholly unsatisfactory state of affairs. Migratory moves can and frequently do occur multiple times in the life history of persons. Therefore, adhering to tradition in census taking, and restricting the collection of migration data to the last move only, effectively means that much of actual ongoing and historical migration processes is simply ignored.

To illustrate the consequences of this, it is useful, here, to draw a parallel with the familiar case of fertility. The event of giving birth can also occur multiple times in the life history of a woman. Now, recall that population censuses are usually conducted only once every ten years. Therefore, maintaining the parallel with migration data collection, suppose that we would collect a data set on fertility only once every ten years. Further, again maintaining the parallel, also suppose that once in those every ten years we would exclusively collect data on the most recent birth for each woman. Thus, if a woman has also given birth earlier to other children, then this is ignored in the data collection procedure. In other words, out of all events (births) that have actually occurred to women over these ten years, we would observe and recognize at the most one event (birth) per woman.

The resulting information on actual true fertility in the population, such as, for instance, the fertility rates and their development by age and over time or the absolute numbers of children born as time progresses, would clearly be highly incomplete and totally unsatisfactory.

Yet, in the case of migration data collection, this is precisely what is routinely recommended and accepted as standard data collection practice.

To this we should then add the empirical fact that, compared with fertility, migration tends to be both a significantly more varied and a much less stable process, not only by age, but especially also over time as well as geographically. Consequently, the voids left in our empirical information are all the more serious and deeply troubling.

It will be clear, therefore, that adhering to the tradition of collecting last move data only is an untenable position.

Instead of submitting to the argument that no change in the essentially many decades old substantive and operational implementation of population census taking is best, and instead of simply maintaining traditional practice as regards migration data collection, national statistical offices cannot but accept the inevitable:

In most countries it is recognized that migration, be it internal or international, is a key issue of such high socio-economic and political importance as to warrant the collection of adequate data through the periodic national population census. This is also the position taken by, for example, the United Nations in its role as the global agency which sets the international standards for population censuses and which coordinates the periodic worldwide population census rounds.

However, as explained, the requirement to collect adequate migration data means that there is no realistic choice but to extend questioning in population censuses beyond merely any last moves only.

In other words, and here we cannot therefore but differ from the aforementioned referees, extending questioning on migration behaviour to earlier migratory moves is not so much a matter of "whether or not", as it is one of "how".

Extending such questioning is clearly a challenging matter, and, given the elementary importance of the task, we tackle this in some detail in chapter 2. There we discuss several complementary and mutually reinforcing *practical operational strategies* to facilitate achieving such more in-depth questioning on past migration behaviour.

It reflects a considered view, based on well over three decades of experience worldwide with the actual practice of census and survey taking, processing and analysis, that the implementation of these strategies is not only desirable but also realistically practicable.

Importantly, these strategies are not merely motivated by a desire to obtain more adequate data sets on migration. They are far more general, and, in optimizing the approach to population census taking, their benefits extend well beyond the collection of data on internal and international migration:

They significantly contribute to improving the *overall information coverage* obtainable through population and housing censuses, as well as to improving the *quality of the resulting data*. And they structurally add both to the *efficiency of the census process*, and to the continued *further professionalization* of national statistical offices in particular in the developing world, while at the same time enhancing the *relevance* of their role as key *information service providers* for evidence-based policy making, planning, monitoring and assessment.

There is no doubt that, especially in many developing countries and emerging economies, this effort to overhaul and modernize traditional census practice will require adequate local capacity building. In order to realize this, it is necessary that global coordinating agencies, such as the United Nations, and international professional bodies, such as the International Statistical Institute (ISI) and the International Union for the Scientific Study of Population (IUSSP), take a lead role here in developing practical guidance and in stimulating initiatives.

6 A Historical Note

Within mainstream demography, population change through relocation (migration) has never enjoyed the same degree of attention as has population change through births (fertility) and through deaths (mortality). In particular in the second half of the 20th century, methods of measuring and estimating mortality and fertility, and associated empirical data collection efforts, have made impressive strides.

In comparison, as regards methods of measurement, the area of migration lagged woefully behind. The only truly notable and influential exception were methods of indirect estimation. Best known among these are the so-called residual methods. These are estimation methods allowing one to infer information on the *net effect* of migration on population development in the complete or partial absence of empirical data on migration itself.

However, *direct methods of actually measuring migration processes* themselves remained underdeveloped. At least partly as a consequence, the matter of *which migration data best to collect*, and *which data sources best to use*, too, remained unresolved and disputed.

Specifically, in terms of data collection, many demographers and government statisticians proved happy with collecting data on numbers of *migrants* (*persons* who experience demographic events), rather than on *migratory moves* (the *events* themselves).

This is fully at odds with the established demographic paradigm, and it results in information lacking in powers of analysis, insight and understanding. That this is so is easy to see when we draw a parallel with fertility: When it would concern fertility, this would equate to collecting data on numbers of *mothers* (persons who experience the events), rather than on *births* (the events themselves).

The new methods of measuring internal and international migration presented in this book date back to the early 1980s. They originated out of dissatisfaction with the overwhelming limits in terms of informational value inherent in indirect estimation, and out of a quest to benefit from the full power of the events-based demographic paradigm.

Interestingly, from an event-perspective, migration is a more general demographic process than mortality or even fertility. Consequently, the resulting *event-based methods of measuring* internal and international migration *apply equally to* such *other demographic processes*.

In fact, the approach taken leads to important new insights into demographic measurement. As it turns out, not only are the resulting new methods of measurement significantly more powerful than the familiar traditional methods of measuring mortality and fertility that have become accepted as today's standard practice in demography. They also clearly demonstrate that these traditional methods are characterized by several avoidable fundamental methodological weaknesses, mathematical flaws and empirical limitations. The matter is discussed in some detail in chapter 1.

This in particular makes it useful to highlight some of the key aspects underlying the development of these new event-based methods of measurement.

These new methods were developed in a deductive manner, as a straightforward logical *operationalization* derived from an abstract and axiomatic formulation of mathematical demographic theory. This mathematical theory, based on stochastic (probabilistic) principles, is general, and it formulates the *development of demographic processes* *along cohort lines* *as time and age progress*, that is, it formulates the *occurrence of demographic events* *in the life histories of cohort members* *as time and age progress*.

This abstract general nature of the underlying mathematical theory also explains why the operational results are equally applicable to other demographic processes, such as mortality and fertility. In fact, for instance both the well-known classical life table and the multistate life table are both specific developments which derive direct from the general theory.

As an introduction, the present textbook does not explicitly develop this mathematical framework. (See also section 3 of this preface, which describes the prerequisites for this book.) It merely highlights and explains some of the elementary theoretical background and principles in a non-mathematical manner.

Instead, this book focuses on the key practical operational implications of the theory, and it is the *resulting* applied methods of measurement of migration which take centre stage:

Chapter 1 deals with the *methods of measuring internal and international migration*, and chapter 2 discusses the associated requirements in terms of *data specifications* and the issue of *data collection*.

However, while the underlying mathematical theory is not an explicit part of the present book, inevitably the deductive nature of this developmental line of thinking is clearly reflected in the presentation. Methodologically, as a consequence, the approach taken differs from the one found in most modern standard textbooks in demography in one important respect: In today's common demographic paradigm, empirical observation and available statistical data tend to be taken as a key point of departure, and elements of mathematical theory are then developed as a framework to structure, interpret and generalize observed patterns and processes. The period life table, based on empirically observed annual mortality rates, and the derived Leslie projection matrix (see also section 2 of this preface) are two typical cases in point. It is an approach which may be characterized as *measurement-based theory development*.

The new methods presented in this book, on the other hand, are based on a reversal of this paradigm: Although this remains outside the scope of this book, abstract mathematical theory is formulated first. And then next, by implication appropriate methods of measurement are logically derived. This approach can be described as *theory-based development of measurement methods*.

Axiomatization and abstract mathematical theory development, theory with is only subsequently confronted with empirical observation, are hallmarks of mature scientific development which in a more general sense began to emerge later in the 19th century. Given its well-defined and quite restricted domain, demography is probably the only social science easily capable of adopting this paradigm. Methodologically, this approach ensures the full internal consistency of the work, and at the same time it allows for its critical assessment in a fully transparent manner. In addition, through its underlying logic, it leads to powerful insights and results in a manner that is both obvious and direct. It is an approach which has proved to be extremely valuable in the development of the methods presented in this book.

As mentioned earlier, these methods were first developed in the 1980s. For readers interested in original historical material we have succeeded in retrieving a copy of a presentation from the mid 1980s [3] which summarizes developments up to that time. It is a succinct statement which, despite progressive insight since, remains valid, comprising many of the key elements of mathematical theory, methodological principle and operational practice. The copy has been restored to a readable state, and it is available on request to interested readers.

Elements in the line of thinking underlying the work presented in this book have much older roots, however. In 1984, while at the University of Manchester, a colleague in the Department drew my attention to the fact that the approach fits in a scientific tradition whose roots can be traced back direct to the 17th century, to the work of Johan de Witt (1625-1672). De Witt was a lawyer and mathematician, and one of the great statesmen of the Dutch Republic. His interest in the republic's finances led him to study life insurance, and thus also mortality [4]. He approached the matter in a rigorous scientific manner for the first time. As a result, De Witt is now widely regarded as the founding father of modern actuarial science.

From our perspective here, several aspects of De Witt's work stand out in particular, since they are also key to the methods discussed in the present book. They include, among other things, his stochastic (probabilistic) mathematical approach, a scientific first not only in demographic analysis; the central position of theoretical mathematical constructs in deductively deriving empirical conclusions; and a firm cohort (longitudinal) perspective, as opposed to the period (cross-sectional) framework that has become so common in applied demographic analysis today.

Clearly, therefore, unaware of it as I may have been at the time, an intellectual debt is owed to De Witt. And so, as a fitting tribute in recognition of the lasting significance of his pioneering scientific achievements for the study of population, it is his image which deservedly graces the front cover of this book.

---

3 Doeve, W L J (1986) *How Do We Measure Migration? The Preferred Migration Questions for the Global 1990 Round of Population Censuses*. Paper presented at the International Conference on Urbanization and Urban Population Problems (ICUUPP), Tianjin, People's Republic of China, October 1987

4 De Witt, Johan (1671); see the references to chapter 1.

In concluding this preface, I wish to thank all those who have commented on earlier versions of the material presented in this book. And, of course, I continue to welcome any further questions, comments and suggestions.

Finally, and most important of all, I am grateful to my wife for her full support in my writing of this book.

William L J Xu-Doeve

E-mail: w.l.j.xu-doeve@anrc-consulting.com

Dedicated to my parents

Jan Willem Doeve (1918-1979)
Gretha Knoester (1923-1965)

SELAMAT JALAN

to all with the courage and enterprise
to leave a familiar environment
in search of better opportunities
and a better future

MAIN TABLE OF CONTENTS

Chapter 1

Chapter 1

The Basic Principles of the Measurement of Migration Using Population Censuses

This chapter was first presented as a paper at the International Conference on Migration and Development (the Fifth Valentey Lecture, Lomonosov Moscow State University), Moscow, 13-15 September 2007.

ABSTRACT

Many countries -- developing and developed alike -- experience very significant difficulties in practice to produce *timely and reliable information* on the *size* of, and on the *trends* in, internal and international *migration flows* and *migrant stocks*.

In particular, from a methodological point of view there still exists considerable ambiguity about (1) the *optimal specifications of migration data* to be collected; about (2) which *data sources* best to use; and about (3) the *most informative methods of measurement* of migrant *flows and stocks* and of *their dynamics over time*.

Consequently, in applied practice, the approaches adopted to the measurement and analysis of migration by National Statistical Offices and other research institutions are often characterized by important *methodological weaknesses*; and the resulting information on migration rates and on migrant flows and stocks is frequently surrounded by a very considerable degree of *empirical uncertainty*.

This first chapter presents new and powerful approaches to demographic measurement which comprehensively address these issues:

It outlines how *methodologically sound*, *mathematically* and *demographically consistent*, *empirically comprehensive* and *fully-detailed* insights into ongoing and historical processes of internal and international migration can be obtained, using conventional duration of residence data.

The chapter centres in particular on the presentation of methods and materials for *direct practical application*, and for the production of relevant and timely information on migration processes to enable and support evidence-based policy making, planning, monitoring and evaluation.

KEYWORDS

Internal Migration, International Migration, Methodology, Demographic Measurement, Population Census

Copyright © 2007, 2008 William L J Xu-Doeve

Table of Contents

1 Background

In recent decades it has become widely recognized that migration -- internal and international -- poses issues which are of major concern for policy making, planning and management in both the public and private sectors in many countries.

For example, in particular in important parts of the developing world, *internal migration* continues to change the face of nations: In many low and lower-middle income countries, the movement of people from rural areas and smaller towns in search of better opportunities and a better future, is contributing significantly to the current massive growth of several of the major towns and cities. Such flows are not necessarily one-way and one-stop. China's "floating population", for example, is highly dynamic and characterized by the importance of staged migration and of return migration (individual migration trajectories). Significant migration flows in the developing world are also the result of less voluntary internal displacements. Familiar causes are adverse natural circumstances such as earthquakes, droughts, floods and hurricanes; development projects such as the construction of dams, large-scale export-oriented commercial agricultural, mining and industrial developments; and conflicts and civil strife.

International migration, too, has risen to the top of the policy agenda, both in the developed world and in the developing world. Rising flows of international migrants, both temporary and permanent, feature more and more prominently in so-called "migration debates". Such debates tend to focus on a variety of often contentious issues, such as the desirability and ability of receiving nations to absorb "foreigners" or "aliens", on topics such as brain drain from sending nations, and on mechanisms to manage flows. At the same time, international migration fills skills gaps and provides cheap labour in wealthier nations, such as in the UK, the USA, or the Gulf States. And it is accompanied by steadily more important financial flows (remittances) back to the countries of origin.

Clearly, all such migratory flows, internal and international, have significant implications for policy and decision making and for management in the public and private sectors in a range of different areas, including, for example:

- the provision of education, of health care, of sanitation, and of other services
- the provision of housing and infrastructure
- the labour market
- skills development and brain drain
- internal and international remittances
- socio-economic and cultural integration of migrant populations
- and so on

Sound evidence-based policy and decision making in such areas demands *timely and reliable information* on the *size* of, on the *nature* of, and on the *trends* in, actual *migration flows* and *migrant stocks*. To date, however, most countries experience very significant difficulties in practice to produce such information. It is useful briefly to highlight some of the factors underlying this.

The study of events affecting the size, structure and development of human populations belongs to the domain of demography. However, one cannot but observe that within the demographic paradigm migration has, de facto, often been considered as quite peripheral to the discipline. Even to date, most *academic and other training programmes* in demography centre predominantly on mortality and fertility.

The 1970s and 1980s did witness the development of what was first called multiregional demography (Rogers, 1975). After further development, and when the generality of the approach was better understood, it became more commonly known as multistate demography (Keyfitz and Caswell, 2005). In population analysis, the central underlying mathematics was first found in much earlier work by Du Pasquier (1912, 1913), but for long this remained unacknowledged. Essentially, multistate demography was a straightforward generalization of multiple decrement life table analysis by allowing not only for decrements (attrition) but also for increments (accrual). This development enabled processes of internal migration (and formally similar "out-in" processes) to be properly accommodated in life table based demographic modelling. The generalization of the classical cohort component approach into multiregional population projections allowing for redistribution through internal migration, was now of course also obvious.

Thus, with the emergence of multistate demography, the *analysis* and *forecasting* of populations experiencing internal migration received something of an impetus. However, the question of the production of adequate data -- the methodologically sound *measurement* of internal and international migration -- was left a poorly developed area. Even today, major *textbooks* in demography either ignore the issue, or at best give it scant and unsatisfactory attention.

For example, one of the major textbooks that does deal with the question of migration data is Siegel and Swanson (2004), formerly better known as Shryock et al (1971). However, the two chapters on internal and international migration together comprise less than one tenth of the text, and the material on the measurement of migration is theoretically, methodologically and operationally outdated.

From various angles, migration does also feature in other academic disciplines, most notably in *geography*, *economics* and *sociology*. Such other disciplines, however, each have their own traditions, paradigms and perspectives. Historically these have proved not usually to encourage the focusing on the theoretical, methodological and operational issues specifically associated with the question of the measurement of migration processes beyond the basic directly observable descriptive level.

Population Accounting and the Measurement of Migration

One notable exception to the involvement of disciplines other than demography in the methodology of measuring migration occurred in geography in the 1970s and 1980s. In direct competition with the demographic multistate approach, Rees and Wilson (1977) put forward a generalized version of population accounting. They disaggegated the well-known classical accounting equation $K_{t+n} = K_t + B_t - D_t$, making it specific by region.

In the above classical version of the equation, K_t denotes the population at some point in time t ; K_{t+n} denotes this population one n-year wide time interval later; and B_t and D_t, respectively, denote the births and the deaths in the intervening n-year period, thus balancing the equation. For a more detailed description of the population growth process, the equation may be specified by age and sex.

Rees and Wilson then simply achieved the regional disaggregation by adding appropriate inmigrant and outmigrant terms for each region into which the population K has been broken down. These terms then represent the persons living in one region at time point t and surviving to live in another region at $t+n$, and they serve to balance the resulting set of regionally-specific equations. In this straightforward way, internal migration was brought into the population accounting framework.

Importantly, while demography focuses on *demographic events* -- such as births, deaths, migratory moves -- and on *demographic rates* -- birth rates, death rates, migration rates --, population accounting centres on *persons*: persons born, persons who die, and migrants. This distinction can have subtle but important consequences for demographic measurement and analysis. Note, for example, that in the case of the analysis of mortality, effectively events and persons can be freely interchanged since the event of death happens only once, at the end of the life time of a person. However, this is not the case in, for instance, the study of migration: migratory events are potentially recurring events in the life history of a person. Here, therefore, events (moves) and persons (migrants) must be distinguished. Further, for example, in demography a birth is a demographic event occurring to a mother at some point in her life history, and the event may recur at other such points. On the other hand, in the accounting framework a person born is simply a newly-emerged population member to be added to the stock K. These distinctions at once underly both the simplicity of demographic accounting principles and their lack of analytical and informative power.

However, although this essentially simple and descriptive multiregional accounting framework compared unfavourably with the multistate demographic approach in terms of its ability to generate insightful information, it was attractive for geographers and regional planners for two reasons. First, it did not require the mathematical skills demanded by multistate demography (even though the multiple-indexed disaggregated accounting equations themselves could be quite cumbersome to read). Second, accounting equations can be interpreted quite straightforwardly as forecasting equations, and so the approach offered a crude but intuitively simple approach to Markov chain multiregional population projections that allow for internal migration.

Most important from our perspective, this multiregional accounting framework led to an obvious choice in respect of the *preferred internal migration data*: surviving persons specified by the place of usual residence one fixed n-year wide time interval -- usually one or five years -- prior to time point $t+n$ (Rees, 1984).

While multiregional population accounting never became the paradigm of choice in demography, its preferred data specification was uncritically embraced by practitioners of multistate demography, and it subsequently found widespread acceptance, especially in the Anglo-Saxon world, in Europe, and in Latin America. By extension, it was also advocated for the measurement of international migration; see, for example Willekens (1994).

This is surprising -- and, in fact, deplorable -- for many reasons. As we just saw, this type of data is incongruous with the demographic paradigm in respect of empirical observation. It does not capture *migratory events* (*moves*); it only captures the surviving *migrants* (*persons*) by origin and by final destination after one or five years, that is, the resulting transitions in discrete time.

Since the demographic events themselves are not observed, it is impossible to evaluate rates -- a cornerstone of demographic analysis -- in a methodologically sound manner. As already shown by Du Pasquier, given demographic rates one can derive the resulting transitions in discrete time using straightforward mathematics. However, the reverse is obviously impossible: given an observed set of interregional transitions in discrete time, there are infinitely many possible sets of events (patterns of migration) that result in the same transitions.

In other words, and perhaps more important from an empirical point of view, transitions *can never reveal the actual ongoing migration processes*. Migration trajectories, including staged and return migration, cannot be ascertained. Frequent, temporary and short-term migration equally are not properly observable. Both the characteristics of, and the trends over time and age in migration behaviour within the one- or five-year period necessarily remain obscure.

Further, the resulting data cannot be adjusted for migration-specific incompleteness. As we shall see later, such incompleteness can render observed migration data for recent migrants totally invalid.

Finally, because true flows and associated stocks are unobserved, any explanatory analysis of migration behaviour is fraught with uncertainty.

There is one organization which, notably since the 1950s, has played a significant and global role in demographic data collection and analysis, namely the *United Nations*. In particular in countries with a more limited statistical capacity, the work and advice of the United Nations are often taken as the leading guidelines in the practice of data collection and in methods of measurement and analysis.

Within the United Nations system, principally two collaborating but separate divisions are involved in setting data collection standards and in methodological development, namely the *Population Division* and the *Statistics Division*; some work also takes place elsewhere in the organization.

Two of the key tasks of the Statistics Division are the promotion of good practice in, and the development of international standards for statistical data collection, including demographic data; and the collating and disseminating of available official national statistics in internationally integrated and comparative systems. The Population Division uses such data in a broad range of interpretive analytical studies, monitoring reports and projections; and it focuses on the improvement of the institutional capabilities of governments in the analysis of national population data, amongst other things through the development and advancement of methodological expertise. From our perspective here, methods of measuring migration, the first task of the Statistics Division and the second activity of the Population Division are of primary interest.

The *UN Population Division* has a long-standing and ongoing tradition of expertise development through the publishing of methodological material in the areas of the estimation and measurement of population structures, mortality and fertility. Remarkably, however, it has only ever once produced a manual on methods of measuring migration: Manual VI (United Nations, 1970).

This short manual, authored principally by KC Zachariah and characterized by an exemplary lucidity of exposition, centres exclusively on internal migration. Today, well over 35 years later, it is no longer up to standard, however: Its principal focus is limited to the *indirect estimation* of the *net result of migration*, rather than on the the direct measurement of actual moves, flows and stocks. In applied practice, such indirect estimation is often based on the comparison of the population size and structure of some urban area, or of some other region, at two points in time. Given information on mortality and fertility in the intervening period, the remaining unexplained difference between the population size and structure at these two time points -- or the *residuals* -- then is *designated as* net migration. Thus, the actual *migration processes themselves remain unobserved*.

Further, interpreting such residuals as representing net migration assumes that the observed population, mortality and fertility data themselves are all free of any error, or that any errors in these data cancel each other out. This is a strong (that is, a restrictive) assumption, which will be violated in many applied empirical cases. However, of course, in applied practice the extent of such violations is difficult if not impossible to ascertain, resulting in unquantifiable uncertainty.

The manual does also briefly discuss some of the most common census and survey questions used in the *direct measurement* of internal migration. However, reflecting the state of the art in the demographic measurement of migration at the time, the manual fails fully to appreciate the merits and demerits of such questions, and, beyond description, it does not exploit the demographic paradigm to derive methods of measurement. Thus, considering the predominant emphasis of Manual VI on residual estimation, its title, "*Methods of Measuring Internal Migration*", is in fact something of a misnomer: In the residual estimation of migration, everything in the equation is measured but migration itself.

While the UN Population Division has a relatively free hand in its choice to author or commission work that advances methodological expertise in the measurement of migration, the the hands of the *UN Statistics Division* in developing international standards for migration data collection and promoting good practice, are much more tied.

As matters work, the Statistics Division's standards, recommendations and guidelines come about in processes of international consultation where ultimately consensus plays a role. Inevitably, many countries are involved, each with different insights, priorities, agendas, policies, practices and traditions. Consequently, there is an important element of accommodation and compromise in key documents dealing with recommendations regarding concepts, methods and practices for migration data collection, including in particular in the two key sets of recommendations United Nations (1997) and United Nations (1998). Xu-Doeve (2006), who analyses the latter two sets of recommendations, concludes that they contain significant methodological weaknesses. As a result, adherence to these guidelines does not guarantee that a country will in fact collect data sets on migration that allow meaningful and maximally informative measurement of ongoing processes of internal and international migration.

We do note that the UN Statistics Division is well aware of these limitations. And, given the constraints of its remit and of its modus operandi, it does actively aim to promote the highest methodological standards and best practice through other initiatives; see, for example, Xu-Doeve (2007).

In summary, however pressing the issues associated with internal and international migration may be for policy and decision makers in many countries, to this day we encounter a scarcity of expertise in respect of even the basic measurement of ongoing migration processes. This applies in particular in nations with a limited statistical capacity, but in many ways it is also a phenomenon encountered in other countries.

In practice, there continues to exist considerable ambiguity about

(1) which *data sources* best to use;
(2) the *optimal specifications of migration data* to be collected;
(3) the *most informative methods of measurement* of *flows* and *stocks* and of *their dynamics over time*.

As discussed by Xu-Doeve (2006), the data sets available in many countries, including in many developed nations, are deficient. Empirical studies are often characterized by important methodological weaknesses. And the resulting information on migration rates and on migrant flows and stocks is frequently surrounded by a very significant degree of empirical uncertainty.

In what follows, we shall not go into detail about the first of these three issues, that of data sources. In most countries, there is actually little choice, in particular if one wishes comprehensively to map out ongoing migration processes affecting the nation. In practice in this case, there are generally only two alternatives, a *continuous population registration system* and the *periodic population census*. Surveys based on random sampling are not normally an option for such comprehensive mapping of migration processes. This is, because in the usual absence of suitable a priori migration-related information that enables efficient stratification strategies, they result in excessive margins of stochastic uncertainty.

Few countries have adequate registration systems, and this usually leaves only the population census as a suitable comprehensive source of data on migration.

Below and in the subsequent sections, we shall discuss and illustrate approaches to demographic measurement which comprehensively address the remaining two issues, those of data specifications and of methods of measurement. Specifically, we shall outline how *mathematically consistent*, *methodologically sound*, *comprehensive* and *fully-detailed* insights into ongoing processes of internal and international migration can be obtained, using no more than conventional <u>duration of residence data</u>.

Duration of residence is a powerful concept demographically: This is, because it uniquely *fixes demographic events* (in this case, *migratory moves*) *in time and in age in the life histories of persons*. As we have seen above, precisely this is central to the demographic paradigm, and it facilitates the deployment of the full power of demographic analysis. However, even without recourse to any demographic analysis, it will already be immediately obvious that, if fully observed, such data *completely capture* past migration behaviour, and thus past *migration processes*, experienced by a population.

Such residence durations can be observed during the entire life span experienced by each person up to the point in time of observation ("full observation"). However, more commonly, such experienced lifetime event histories are not observed completely ("partial observation"). In particular in population censuses, observation is often limited to *the most recent migratory move* only.

Clearly, such restricted observation directly rules out the possibility to measure important issues such as staged migration, including its trajectories and timings in the life histories of persons. Therefore, if one wishes to answer many of the common deeper questions about ongoing migration processes and gain valuable and detailed additional insights, then extending observation to include *<u>at least</u> the two most recent events* (*moves*) is, in fact, necessary.

As we shall see, measurement results can include, for example, detailed information on:

(1) migration *rates*, migrant *flows* and *stocks* and their *dynamics over time and age*;
(2) migration *trajectories*, including circular and return migration;
(3) *temporary* and *short-term* versus *long-term* migration;
(4) *frequent* migration; and
(5) *estimates of* and *adjustments for* the often highly *incomplete enumeration* or *registration* of migrants.

The results also lead in a transparent manner to approaches for

(6) more *in-depth analyses*; and for
(7) *explanatory studies* of migration behaviour.

In the next sections we shall set out the basic principles of the demographic measurement of migration. The emphasis will be on *applied measurement* based on *population census data* in developing countries and in emerging and middle-income economies.

These are methods and materials which are elementary for even the most basic production of relevant and timely information on migration processes with a view to enabling and supporting evidence-based policy making, planning, monitoring and evaluation in policy areas such as those mentioned earlier.

In focusing on population census data, a key objective is to provide sound *guidelines* in preparation for migration data collection and for the measurement and analysis of ongoing migration processes in the *2010 global round of population censuses*.

The illustrative data presented are on *internal migration*. However, the methods discussed are general. They easily and directly extend to

(1) the measurement of *international migration*; to
(2) the use of *alternative data sources*, such as *continuous population registers*; and to
(3) *alternative demographic processes*, such as *mortality* and *fertility*.

In fact, we shall make frequent reference to the measurement of mortality. There are two main reasons for this. First, most demographers and statisticians are well acquainted with mortality analysis. Second, as we shall see, life tables, a familiar tool used in the study of mortality, derive from the very mathematical basis that underlies demographic measurement.

However, our approach will be non-traditional in one important way. Commonly, demographers proceed from observed mortality rates to derive probabilities of death and survival. Models, such as life tables and cohort component projection models, are then derived from these probabilities. The approach is captured tellingly in the title of the well-known paper by Keyfitz (1970): "*Finding Probabilities from Observed Rates or How to Make a Life Table*". It is an approach which may be called the development of *measurement-based theory*.

Our approach is essentially opposite. We shall begin by the development of demography's mathematical theory, based on a number of elementary probabilistic axiomatic postulates. (In our presentation here, we shall limit ourselves to no more than a simple and non-technical sketch.) From the results, we shall then logically derive the specification of the necessary data, as well as methods of measurement. Thus, our approach is one of *theory-based measurement*. Following this line of reasoning, it will become clear that the common traditional approach of measurement-based theory development is *fundamentally flawed* in important respects, and that it *lacks the analytical and empirical power* inherent in theory-based measurement.

To conclude this section, let us briefly outline the organization of the remaining sections. First, we shall set out our basic conceptual and operational framework. Then, in section 3, we shall briefly highlight the key aspects of demographic theory that are essential to understanding the principles of demographic measurement. Next, in section 4, we present these basic measurement principles, and we describe how to operationalize them in applied contexts. The next section then illustrates the measurement procedure in the case where we have population census data. Section 6, finally, presents a number of key conclusions.

2 Conceptual and Operational Framework

In this section we shall briefly discuss a number of elementary conceptual and operational notions. This serves to provide the basic vocabulary as well as the analytical framework for our subsequent discussion of demographic theory and of the principles of demographic measurement resulting from that theory.

Migration is the *change of the place of usual residence of a person from one migration defining area (MDA) to another*. When dealing with *international migration*, that is, migration across national boundaries, then the *migration defining areas (MDAs)* are often defined as *countries of usual residence*. However, it is a restrictive choice to do so, since it limits the precision of the geographical specification (the geographical resolution). It is not a necessary choice, nor may it always be a useful choice.

In the case of *internal migration*, places of usual residence are classified into some set of subnational geographical regions. Clearly, any given classification rarely suits all analytical purposes. It is therefore not a good choice to build in some a priori classification into the data collection process. It is recommended instead to use *geo-referencing* in collecting data. This then allows data users subsequently to measure and analyse migration processes using migration defining areas (MDAs) that can be flexibly tailored to the specific issue under investigation.

The question of what in fact constitutes *usual residence* has proven to be a contentious one. In practice, we often see prejudged and de iure approaches which severely limit the ability of analysts of the resulting data to obtain comprehensive, detailed and meaningful insights (Xu-Doeve, 2006).

Operationally, therefore, the recommendation is to consider *every change of residence as a change of usual residence*, *unless* it is both *short*, that is, under one month, and at the same time for a *purpose of stay* which clearly suggests that pre-existing usual residence has not changed.

The recommended exhaustive list of purposes of stay that disqualify such short residence as usual residence, comprises: recreation, tourism and holidays; business travel; temporary visits to friends and relatives; temporary medical treatment; and religious pilgrimage.

Only such a definition leaves the analyst with the most flexibility to investigate all aspects of the migration process; including, for example, key issues such as recent migration, frequent migration, short-term and temporary migration.

Next, let us briefly explore the pivotal role of demography in studying migration. We shall do so in somewhat formal abstract terms. The reason for this is, that it helps to highlight and underscore the generality of the approach that we shall take.

Demography is the science that studies *how populations are affected by the occurrence of demographic events in the course of the life histories of population members*. Populations tend to be internally heterogeneous. Demography handles such heterogeneity commonly by subdividing a population into internally more homogeneous *cohorts*. Well-known traditional examples are cohorts specified by age and sex. In the case of migration, such cohorts at some given point in time and age are further specified by migration defining area (MDA). As always, cohorts are studied individually; and, as appropriate, they may subsequently be reassembled into populations for an overall picture.

Migration is, of course, a key class of events within the domain of demography. A change in the place of usual residence of a person is an event occurring in the life history of this person. If it involves the crossing of a boundary between any two MDAs (or, somewhat more precisely, if the MDA of origin and the MDA of destination are not identical), then the event is also a demographic event, and then it is called a (*migratory*) *move*.

The event is, in fact, demographic in that case, because it affects the distribution by MDA of a cohort that has been specified by MDA at some given point in time and age. (Henceforth, we shall not use the term event unless it is also a demographic event.) Specifically, for example, out- or emigration reduces the number of members of a given cohort in the MDA of origin, and it increases the number of its members in the MDA of destination. Overall, considering all cohorts together, the result is a *changing size and structure of the population of each MDA* through *internal* or *international population redistribution*.

It is useful to put matters in a broader demographic perspective. Other well-known demographic events are the event of giving birth and the event of dying. *Formally*, from a demographic perspective, *processes of birth, death and migration are similar*.

Specifically, each such process relates to a specific *status* (or *condition*) of cohort members: the life status, the fertility status, and the migration status, respectively. Each status is classified into an exhaustive set of non-overlapping *demographic states*. A demographic state is a *categorical variable* that characterizes the position (or, methodologically more precisely, the *value*) taken by individual cohort members within the corresponding status. The study of demography then focuses on *demographic events*, defined as <u>*changes*</u> <u>*in the*</u> <u>*state-value*</u> *of cohort members* <u>*over age and time*</u>.

To make these abstract concepts more concrete, let us give a familiar example. In the case of mortality, the life status of cohort members is the key area of study. The life status can be exhaustively classified into two non-overlapping states, the alive state and the death state. The events of interest occurring to cohort members are defined as a *change in the value* of the categorical variable state that is *taken by cohort members*. Thus, for instance, the event of dying is formally described as a person's state variable changing from the value alive to the value not-alive. The events in question (formally, that is, the value changes) are then observed *as a function of time as cohort members progressively age*.

Empirically, in fact, changes from the death state to the alive state have turned out to be uncommon, and this class of events is generally ignored by demographers. Historically, this has always simplified the study of mortality. As a consequence, however, it has also obscured the generality of the approach.

In addition, in the study of mortality, the event of dying is taken as occurring only once in a lifetime. Consequently, the number of events of dying and the number of deaths among cohort members over age and time are identical. Again historically, this has led to an indifferent approach where <u>*persons*</u> and <u>*events*</u> tend to be *poorly differentiated*. This has further contributed to obscuring generality.

In the study of migration, the migration status takes centre stage. This status is classified in demographic states defined in terms of some exhaustive set of non-overlapping migration defining areas (MDAs).

As we just saw, in the case of migration, the demographic event of central interest is defined as a change in the place of usual residence of a cohort member from one MDA to another. Above, we also already noted the common shorthand expression for such an event: a (migratory) *move*. Collectively, moves specified by direction, that is, specified by MDA of origin and by MDA of destination, are called *flows*.

MDAs can be defined as geographical subsets of one country or of more countries, and it is important to stress that from a formal analytical point of view there is *no difference between internal migration and international migration*.

Empirically, migration differs from mortality in two important ways: In principle, the events can take place in any direction between all states (MDAs) distinguished; and events can occur more than once in the life history of a cohort member. Methodologically, this has two consequences.

Consider a cohort specified by the common demographic criteria such as age at some given point in time and by sex, by MDA at a (usually, but not necessarily, the same) given point in time, and possibly also by any other covariates of interest. Then, first, it is important to *differentiate demographic events by trajectory*, that is, by the path taken across the set of MDAs.

Second, unlike in the case of mortality, it is essential to *distinguish between events (moves) and persons (migrants)*. While, as we shall see, in proper cohort analysis the numbers of persons (migrants) involved can be deduced unambiguously from the number of events measured, it will be obvious that the reverse does not hold true:

Because migration is a *potentially recurring event* in the life history of each person in a cohort, the knowledge of the number of migrants in a cohort does not, generally, allow any inferences about the *number* of demographic events (moves) that have occurred, nor about their *timing in the individual life histories* of the cohort members involved, nor about the *trajectories* involved.

Crucially, therefore, *counting persons (migrants) as the point of departure for the measurement of migration will necessarily leave the process of migration experienced by a cohort, and thus by a population, obscure*. Clearly, also, from a methodological perspective, it runs counter to the demographic paradigm.

Yet, in the actual practice of the measurement of migration, counting persons (migrants) is an extremely common approach. Well-known examples are: *persons classified by place of birth*; *persons classified by place of usual residence at some fixed point in time in the past*; *persons classified by nationality* or *by citizenship*.

In all such cases, the "migrant status" is inferred from the measured variable (the place of birth, the place of usual residence at the given earlier date, the citizenship or nationality, respectively), given the current place of usual residence.

Measuring such variables is an inevitable necessity when counting persons, because *the demographic events themselves are not measured*. And, importantly, as a consequence, *the underlying migration processes themselves are never observed, nor can they ever be inferred*.

As regards the measurement and analysis of migration, there is one further methodological point of interest. Empirically, individual migration behaviour is a phenomenon that tends to be strongly related to a person's life cycle: Often we see relatively high intensities around ages [15, 25) [1]; moderately high intensities around the retirement age; and a family life cycle-dependent relatively high intensity at the youngest ages. At the same time, furthermore, factors affecting migration behaviour are very much related to socio-economic, cultural and sometimes political conditions, conditions which are decidedly less than stable over time.

This reinforces the need to emphasize the core paradigm of demography as a science that essentially centres on the analysis of cohort behaviour, that is, a science concerned with the s*tudy of demographic events in the life histories of cohorts*. While this may seem obvious after our preceding discussion, applied practice often abandons this principle.

The history of demography as a science owes much to actuarial studies. Although the origins of actuarial science, laid in the 17th century by De Witt (1671), are firmly centred on cohort analysis, this approach has subsequently been corrupted:

It has become common practice to assemble empirical data from the range of cohorts from young to old at a single given point in time in so-called *period life tables*, and to present this *cross-sectional mixture of demographic behaviours* as the "actual" behaviour of a single, albeit -- formally -- synthetic, cohort.

In the past, such period life tables proved commercially attractive in the life assurance business. Since the late 19th and the early 20th centuries, due to gradually improving sanitation and health care, the industry saw itself structurally faced with slowly declining age-specific mortality intensities over the actual lifetimes of insured lives. It turned out that under these conditions period life tables often tended to exaggerate the actual lifetime risk for lives that were young at the time of entering into a contract. This was interesting, since it implicitly allowed the setting of safely-conservative, and de facto somewhat inflated, premiums.

And at the same time the period life table provided a practical -- and, at least to

1 In the common mathematical notation for intervals, a square bracket denotes "inclusive of the boundary value itself", and a round bracket denotes "exclusive of the boundary value itself". Thus, for example, an age interval specified as [10, 15) denotes the interval that runs from 10 (inclusive) through to 15 (exclusive). This notation is preferable to the equivalent notation 10 - 14 traditionally used by many demographers, because it avoids the ambiguity inherent in this traditional notation.

the premium-paying layman, an impressive -- semblance of certainty in the face of uncertain lifetime risk intensities in the future. It is a certainty which, however, cannot be verified, let alone be falsified, and which therefore has no empirical basis in science.

In fact, from a methodological point of view there are, of course, *no grounds in general which could justify the assumption that such a synthetic cohort from a period life table would represent the behaviour of any actual empirical cohort over its lifetime*. Yet, in effect, the period approach to life table construction has been adopted widely as an established empirical approach in the formation of demographers and in applied demographic practice.

From our perspective here, the measurement of migration, this is an especially important issue. As we shall see, the key underlying principles encountered in life table construction are elementary to the measurement of all demographic events and processes, not only in the study of mortality but equally in the study of, for example, migration and fertility.

While trends over time in mortality intensities by age may be relatively slow, somewhat mitigating the error committed by taking a synthetic cohort as representative of empirical cohorts, this does not hold true in the case of fertility, and even less so in the case of migration. Hence, *adopting a rigorous cohort perspective in the methods of measuring migration is essential*.

With this final observation, we conclude our discussion of the conceptual and operational framework for the measurement of migration. In the next section, we shall very briefly explore the key elements of demographic theory that underlie general demographic measurement methodology.

This is also useful, since it places the familiar concept of the life table in a proper perspective, and in so doing it reinforces the connection between demographic measurement and the principles of life table construction. As a result, the next sections present the basic principles of the measurement of migration in a context familiar to all demographers.

3 Demographic Theory

The power of demography rests on a quite elementary and simple unifying mathematical theory. While this theory is rarely made fully explicit from first principles, its postulates and its derived results underlie both classical and multistate demography.

Demography's mathematical theory also leads directly and logically to the methodological principles of demographic measurement. Therefore, in this section we shall briefly summarize the elementary aspects of this theoretical construct insofar as this is relevant for demographic measurement. In order to make this section as accessible and non-technical as possible, we shall avoid any mathematical proofs. Instead, we shall present the key results, emphasizing interpretation and significance.

3.1 Preliminaries

Consider a state-specific cohort, that is, a cohort specified not only by the usual variables age and sex and by any other variables of interest, but also by demographic state at some given point in time and age. During its entire lifetime, this cohort is subject to *demographic forces*.

A demographic force is denoted in general by μ. Such forces are elementary concepts that are unspecified except that they be non-negative real-valued continuous functions of time and age. They are taken as the *generators of demographic events*.

The term force is, in fact, appropriate: demographic forces have both *magnitude* and *direction*. The magnitude of a demographic force determines the probability that a demographic event occurs: the higher the magnitude of the force, then the greater will be this probability. Further, since any demographic event corresponds to a change from one given state to another, demographic forces also have direction, defined on the state space (the set of categorical states).

Thus, for example, the force of mortality to which a cohort in the alive state is subject, is the generator of the events of dying experienced by the cohort as age and time progress. Similarly, the force of migration generates the migratory events (moves) experienced over age and time by a cohort.

As mentioned, demographic forces μ are continuous functions of time and age. Time and age are two variables that run in parallel on the same scale. They can be used interchangeably, as long as the cohort has been properly specified by age at some given point in time. From a theoretical point of view it proves convenient (but not necessary) to use time variable t in preference to age variable x. Therefore, if we wish to highlight the functional nature of a demographic force, we shall usually write $\mu(t)$.

As an elementary principle, demographic forces are considered *cohort specific*. Thus, for any demographic force $\mu(t)$ it is necessary to be specific in respect of the cohort to which this function refers. By implication, if a cohort has also been specified by any additional attributes, such as sex, then the demographic forces acting on this cohort are also considered specific by these attributes. Further, it is necessary to *specify the direction* of demographic forces. In the case of migration, this entails specifying forces of migration by trajectory.

It helps avoiding ambiguity if, in noting down demographic forces, they are properly indexed, clearly indicating all applicable specifications. However, such full specification can all too easily make for rather reader-hostile notation, turning the μ's into Christmas trees well-decorated with indices. Fortunately, in practice much can usually be inferred from the context of application; and for a cohort of a given age at the benchmark point in time, usually conveniently set as $t = 0$, a time variable t as the function's argument often suffices.

3.2 The Three Axiomatic Postulates

Given these preliminaries, demographic theory construction then begins by formulating *three axiomatic postulates*. All theoretical results presented later in this section are derived mathematically from no more than merely these three postulates. Since demographic measurement is in turn directly derived from demographic theory, the postulates are fundamental to demographic measurement, as well. The postulates are, respectively:

$$P_1(t, t+\Delta t) = \mu(t)\cdot\Delta t + o(\Delta t) \tag{1}$$

$$P_{n>1}(t, t+\Delta t) = o(\Delta t) \tag{2}$$

$$P_{n_2\,|\,n_1}(t_2, t_2+a_2; t_1, t_1+a_1) = P_{n_2}(t_2, t_2+a_2)\ , \quad \text{whenever } [t_1, t_1+a_1) \cap [t_2, t_2+a_2) = \emptyset \tag{3}$$

In these postulates, the function P denotes the probability that a cohort member experiences the indexed *number of events* on the time interval specified as this function's argument (that is, in its brackets); Δt denotes a short time interval; and the function $o(\Delta t)$ vanishes asymptotically (that is, if Δt approaches zero).

All subsequent theory derives from these three postulates. Let us briefly explore what, in fact, these postulates express.

Formulated simply, postulate (1) states that the probability to experience one event (to make a single move in the case of migration) on the Δt-wide time interval from t to $t+\Delta t$ is approximately equal to the magnitude of the demographic force that operates at the time (and age), multiplied by the width Δt of this time interval (the length of the period of operation of the force, or the duration of the exposure to the force). Further, since $o(\Delta t)$ vanishes as Δt approaches zero, this first postulate also states that this approximation becomes better and better, the shorter the length Δt of the time (and age) interval considered.

Postulate (2) effectively implies that demographic events are mutually exclusive if the time interval Δt approaches zero. For example, in the case of migration, it excludes that two moves are made simultaneously.

Postulate (3), finally, states that the numbers of events, n_1 and n_2, respectively, that occur in any of two non-overlapping or disjoint time intervals $[t_1, t_1+a_1)$ and $[t_2, t_2+a_2)$, are stochastically independent.

These three postulates cast the *frequency of the occurrence over time (and age) of the demographic event in question* among cohort members as a *stochastic Poisson process*, one of the well-known counting processes of numbers of events occurring over time.

For any theory resulting from these three postulates to be valid, the postulates themselves must, of course, be empirically plausible. Let us briefly consider this issue.

Clearly, from an empirical demographic point of view, the first two postulates are entirely reasonable. Postulate (3), however, is restrictive since it describes the demographic process as non-hereditary or memory-less.

Postulate (3) implies, for example, that information on cohort behaviour on past intervals of time and age does not contribute to improving predictions made about behaviour on any subsequent intervals. Empirically, this is, of course, not in general reasonable.

For instance, it is generally recognized that persons with a history of intensive migration behaviour, such as frequent migrants, are more likely to make another move later in life than persons with little past migration experience. Or, to give another example, the phenomenon of return migration indicates the selective influence that a past move can have on future migratory behaviour.

One approach to meeting this objection then is weakening (relaxing) this postulate. Effectively, this means replacing the above postulate (3) by one or more alternative postulates that do result in recognizing and incorporating hereditary behaviour (memory).

However, the consequence of this approach is that it inevitably leads to a more complex mathematical formulation of demographic theory. In practice, the negative result of realistically incorporating memory is that the mathematics underlying demographic analysis will very quickly become so complex that common data sets will all too soon prove insufficiently informative actually to calibrate the theory.

In other words, the dilemma is that while the theory would gain in terms of *face validity* by modifying postulate (3), *operationalization* of the resulting theory would quickly become impossible. (It is also interesting to note that most of modern demographic analysis in fact tacitly assumes postulate (3). So, a reformulation of this postulate would render demographic analysis as we know it today obsolete.)

The alternative is to do what generations of demographers have been doing as a matter of course: That is, to attempt to eliminate heterogeneity in demographic behaviour -- such as, here, in historical behaviour -- by subdividing (partitioning) heterogeneous cohorts into internally more homogeneous subcohorts. Statisticians refer to this as *controlling for* such historical behaviour.

In the case of migration, this may, for example, mean distinguishing between cohorts not only on the basis of age, of sex, and of MDA at some benchmark time point, but also on the basis of relevant *characteristics of* one or more *earlier migration trajectories* experienced. Obvious characteristics to consider include the timing and the direction of all, or of selected, migratory events in the life history of persons prior to the benchmark point in time.

Thus, consider, for instance, a study of the migration from India to the United Arab Emirates of, say, a female cohort aged [20, 25) at benchmark time $t = 0$. Then heterogeneity in respect of historical migration behaviour may be reduced by differentiating within this cohort on the basis of the characteristics of earlier India-UAE and/or other migration behaviour experienced. Such subcohorts may then be analysed separately, significantly mitigating the restrictive nature of the third postulate.

We note that, for more approximate partitioning, one or more suitable proxy variables may be used instead. However, in practice it may well be difficult to select suitable proxy variables. The success of eliminating relevant heterogeneity in historical migration behaviour of persons through the use of proxy variables is very much dependent on the association between such proxies and experienced migration histories. The true quality of the propositioned associations is, however, difficult to establish unless reliable information is also available on the migration histories actually experienced. If the latter information is available, then it may in fact well be preferable to use this information direct in the partitioning (that is, the heterogeneity eliminating) procedure.

Next, let us consider the significance of our postulates from the perspective of demographic measurement. We shall do so by presenting a number of fundamental results that directly lead up to methods of demographic measurement.

3.3 The Probability to Experience *n* Events on any Time and Age Interval

From the three postulates it follows directly that the probability $P_n(t)$ of exactly n events, such as n migratory moves, $n = 0, 1, 2, 3, \ldots$, occurring to a cohort member during the time interval $[0, t)$ is given by

$$P_n(t) = \frac{\left(\int_0^t \mu(u)du\right)^n}{n!} \exp\left(-\int_0^t \mu(u)du\right) , \tag{4}$$

a result that is elementary to prove by mathematical induction. Equation (4) is a fundamental result in demography. Statisticians will recognize it as a generalized Poisson distribution. In the study of migration, it is, for instance, of direct use in the measurement and analysis of multiple moves. It is, for example, also a key equation in the collection of migration data (and of other types of data) through capture-recapture random sampling designs.

However, if we limit ourselves to the basic measurement of demographic processes, such as mortality, fertility and migration, then the special case of equation (4) where $n = 0$ is of most immediate interest. Therefore, rather than exploring equation (4) in more depth here, we shall only focus on this special case that zero events are experienced.

3.4 The Probability to Experience No Events on any Time and Age Interval

We recall that μ is a non-negative real-valued function. So, clearly, for any non-negative value of t the integrals in equation (4) evaluate to a non-negative real value, and this value raised to the power 0 equals unity. Further, recall that $0! = 1$.

Therefore, the probability $P_0(t)$ of exactly zero events occurring to a cohort member on the time interval $[0, t)$ equals

$$P_0(t) = \exp(-\int_0^t \mu(u)du) \ . \qquad (5)$$

In interpreting equation (5), recall that $\exp(x)$ denotes the exponential function e^x where the constant e is the base of the natural logarithms. Its value is approximately equal to 2.71828. As to the negative exponent, this may simply be taken as notational convenience avoiding fractional notation; for example, $e^{-x} = 1/e^x$.

Finally, for the benefit of readers unfamiliar with calculus, let us briefly explain the meaning of the definite integral $\int_0^t \mu(u)du$ of the demographic force function $\mu(u)$ over u from 0 to t. Readers familiar with this integral may opt to skip the next few paragraphs set in smaller type.

> This integral can be interpreted as follows:
>
> Consider that we exhaustively subdivide the interval $[0, t)$ in a number of non-overlapping subintervals. Let us make each of the subintervals small enough for the magnitude of the demographic force $\mu(t)$ that operates on the cohort to be *approximately constant* on that subinterval. Individually for each subinterval we then multiply the magnitude of the demographic force on that subinterval by the length of the subinterval. Finally, we *sum* the resulting products. Now, the shorter and shorter we choose the length of each subinterval, then the closer and closer to the truth will be the assumption of approximate constancy of $\mu(t)$ on that subinterval. The integral then is the limiting case where we let the subinterval lengths

asymptotically approach zero.

Clearly, therefore, this integral is a *cumulative function of the demographic force* $\mu(t)$ over time (and age). Effectively it represents the *magnitude* of the demographic force $\mu(t)$ that operates on the cohort, *multiplied by* the length of the *duration of exposure* to the operation of this force, while allowing for the fact that the force $\mu(t)$ is not (or, more precisely, not necessarily) constant on the interval $[0, t)$ under consideration. (The integral is in fact a monotone increasing function of t because $\mu(t)$ is a non-negative real-valued function.)

Put more informally, this integral represents the sum-total of the force of mortality to which a cohort member has been subject from benchmark point in time 0 to point in time t.

We shall return to equation (5) below. However, before doing so, we first state another result that follows direct from the three postulates.

3.5 Demographic Forces and Instantaneous Demographic Rates

A key result under these three postulates is that, locally around any value of t, the magnitude of the demographic force $\mu(t)$ asymptotically approaches the instantaneous demographic rate (or the intensity of the demographic process). The proof of this theorem is again elementary and is left to the reader.

Thus, we also have an unambiguous mathematical and physical interpretation of the hitherto unspecified concept of the magnitude of a demographic force: In the case of migration, for example, we have that the magnitude of the *force of migration at any point in time t* is *equal to* the *instantaneous migration rate*, or the *migration intensity*, *at that point in time t*.

In practice, therefore, we can often use these two terms of demographic force and of instantaneous rate or intensity interchangeably, and henceforth we shall do so.

Further, since in our stochastic framework cohort members are considered at risk of demographic events, the instantaneous rate is also sometime called the *hazard rate*. Appropriately, therefore, the integral in equation (5) is also called the *integrated hazard* or the *cumulative hazard*.

3.6 The Number of Cohort Members that have Not Experienced the Event

Building on equation (5), it also follows that if we have a cohort of size $K(0)$ at benchmark point in time $t = 0$, then the number of members $K_0(t)$ that will have experienced *zero demographic events* by point in time t is given by

$$K_0(t) = K(0) \cdot P_0(t) = K(0) \exp(-\int_0^t \mu(u) du) \ . \qquad (6)$$

Clearly, $K(0)$ is merely a shorthand notation for $K_0(0)$.

Readers familiar with mathematical demography will immediately recognize equation (6) from life table analysis, stable population analysis and other population modelling.

In, for instance, life table analysis equation (6) is encountered as the equation that describes *survival* (zero events of death having occurred among cohort members) from an arbitrary exact starting age x_1 to exact age x_1+t. Commonly in a life table the symbol ℓ (for lives) is used instead of K; the exact age for the start of the analysis x_1 is often taken as zero; and the width of the time interval is usually expressed in terms of exact age x (rather than time t), allowing one -- reckoning from exact age zero -- to write ℓ_x for $K_0(t)$. Clearly, then, $K(0) = \ell_0$, a number also called the radix in life table construction.

Note, incidentally, that postulate (3) implies that we actually are free to set our benchmark point in time $t = 0$, the departure time point for our analysis, to correspond with any exact age of the cohort under study.

Importantly, equation (6) is a general result, in that it applies not only to mortality but to all classes of demographic events, including migration and fertility. Further, it is *a key equation for demographic measurement*.

In preparing the ground for our discussion of demographic measurement, let us briefly remind ourselves of the approach commonly adopted in demography. Let us again take the familiar case of mortality by way of example.

Measurement then consists of evaluating annual age-specific mortality rates M_x by dividing the number of observed events of death in a year by the population at risk during that year. Next, theory construction begins by evaluating the numbers of survivors ℓ_x for all ages from an arbitrary radix ℓ_0 in an annual piecewise approach, using (usually some approximate form of) equation (6) and assuming that $M_x = \mu_x$. Refinements are sometimes applied, but they do not materially alter the principle of the approach.

In the next section, we shall, however, see that this approach is problematic. Specifically, this traditional approach to measurement and theory construction will be shown to be

(1) *mathematically* and *methodologically flawed*;
(2) *difficult to generalize* to classes of *events that are potentially recurring* in the life history of cohort members, such as migratory moves and births;
(3) unnecessarily *limited in the scope and depth of the resulting information* on the demographic processes that have actually taken place.

These objections can, in fact, all be met by reversing the approach, that is, by taking the theoretical results such as those presented above as the point of departure, instead of any empirically observed annual rates M_x.

Given these theoretical results, elementary demographic measurement, for instance, then begins by empirically observing appropriate sets of (t, $K_0(t)$) data points for a given cohort. In the case of mortality, this would simply be a set of observed data on numbers of surviving cohort members as time (and age) progress.

As we can see from equation (6), whatever the demographic process -- be it mortality, fertility or migration --, the corresponding *demographic force function* (or, equivalently, the instantaneous demographic rate function) $\mu(t)$ is *the sole generator of such data points*.

The essence of demographic measurement then simply is the recovery of the $\mu(t)$ function, given a set of cohort data points (t, $K_0(t)$). These are the issues that we shall explore next.

4 Demographic Measurement

In concluding the previous section, we outlined the essence of the elementary principles of demographic measurement, and we contrasted this with common traditional approaches. In the present section, we shall explore the issues involved in greater depth. Further, we shall discuss how to operationalize the approach using common standard methodological procedures.

As we can see from equation (4), all demographic processes are fully characterized if and only if the demographic force $\mu(t)$ is known. Thus, *demographic measurement* -- be it of mortality, fertility, migration or whatever other demographic process -- *fundamentally centres on finding a calibrated expression for* $\mu(t)$.

The standard "textbook approach" to empirically observing $\mu(t)$ that is almost universally instilled in demographers involves a number of steps. By way of example we very briefly discuss the well-known case of mortality and the construction of classical life tables, omitting some refinements which are sometimes made. For further details, the reader is referred to common textbooks, such as for instance Keyfitz (1977), Pressat (1983), Shryock et al (1971), or Siegel and Swanson (2004).

In the first step, *empirical annual* mortality *rates* M_x are obtained for a succession of cohorts throughout the age range in a given year. Commonly, these rates are each evaluated by dividing the observed number of *occurrences* of the event of death among cohort members by some measure of the *exposure*. The exposure is the total number of person-years during which the cohort members at risk of dying have been exposed to this risk.

Two commonly-used approximations of the exposure are: the mid-year cohort size; and the average of the year-start and year-end cohort sizes.

Not seldom, however, are cohorts improperly identified here: Frequently we see that the exposure of each cohort itself is incompletely assessed, while instead a part of the exposure of the two adjoining cohorts is included. Also, in applied practice, the observed occurrences of deaths are sometimes incorrectly allocated to cohorts. These issues of the proper allocation of events to cohorts and of the proper identification of cohorts at risk is discussed in some depth by, for instance, Pressat (1983).

In the next step, the assumption is made that the instantaneous cohort mortality rates $\mu(t)$ are *piecewise stationary*: specifically, they are assumed to be *constant* on each successive age interval. In other words, by way of approximation it is assumed that locally, that is for successive intervals of time and age, it is true that $\mu(t) = \mu$. Under this assumption, equation (5) simplifies to

$$P_0(t) = e^{-\mu \cdot t} \ . \tag{5a}$$

In applied practice, additional approximations are common: For each individual age interval, the RHS of equation (5a) is often approximated by the well-known

$$1 - \{ \mu / (1 + ½ \mu) \} \ = \ (1 - ½ \mu) / (1 + ½ \mu), \tag{5b}$$

usually with refinements for the youngest and oldest age intervals.

Equation (5b) implicitly assumes linear exposure, and its approximation of equation (5a) is notoriously poor. Of course, with well-established tools from numerical analysis and with today's calculating power, such poor approximations are also unnecessary and obsolete.

Next, in the penultimate step, the empirically observed annual rates M_x are assumed equal to the stationary instantaneous rates μ. That is, it is assumed that $M_x = \mu$.

Finally, the *synthetic cohort* ℓ_x is constructed. This is done using a piecewise approach, commonly by multiplying successive survival probabilities that result from equation (5a), or from its approximation, and substituting M_x for μ.

This traditional approach to measurement and analysis is *problematic*, both *mathematically* and *methodologically*. In addition, the approach *does not maximize the informative power inherent in demographic theory and in data sets commonly available*.

Some of the problems we have already discussed. Apart from repeated and suboptimal approximations, this includes in particular the *treatment of period (cross-sectional) data as cohort data*.

It may well be true that in actual practical applications the resulting cohort is correctly characterized as synthetic. However, as discussed in section 2, above, this implies that there is little if any general methodological justification in interpreting the cohort-related findings, such as life expectancies, as information pertaining to some actual empirical population.

Yet, as an applied science, demography's objective is to produce just such empirical information relating to real populations. And all too often in applied practice in the presentation of statistical results, we see that the caution not to interpret cross-sectional data as if they were proper empirical cohort data, remains unheeded. This holds true not only in the study of mortality, but equally in the study of fertility, where statistics such as the total fertility rate (TFR) are commonly based on cross-sectional data.

The traditional approach to demographic measurement, starting with the observation of empirical annual rates of the type M_x, is yet more problematic for a number of further methodological and operational reasons. These are quite fundamental issues that, however, often receive little attention. They relate both to the evaluation of the *numerator* and to the evaluation of the *denominator* of such empirically observed demographic rates.

The *numerator* of such rates is commonly established by *observing* the *number of events occurring* to cohort members during a given year. This number is often referred to simply as the ***occurrence***. Here, the key problems are with data sources, with the potentially recurring nature of demographic events, and with incomplete observation.

Empirically observing the occurrence of course requires the availability of a *data source* in which such events are actually recorded. Commonly, for mortality and fertility this source is a continuous vital registration system or some form of longitudinal household survey. Outside the Western world, however, there are few countries that maintain effective vital registration systems. Further, the coverage of household surveys, where they exist, is rarely nationwide and not unusually time limited.

This, of course, explains the ongoing reliance in many developing countries on indirect methods of mortality and fertility estimation. It is, however, important to note that many of the key indirect methods used here -- including commonly-used model life tables and model fertility schedules -- ultimately derive from sets of reliably observed *period* mortality and fertility *data* that have been *applied to* *synthetic cohorts*. In addition, the validity of such model schedules in any context of empirical application is, of course, impossible to verify, let alone to falsify: The very reason for their use is the absence or inadequacy of reliable empirical data. Effectively, this makes the application of model schedules as substitutes for inadequate or unavailable empirical data rather much an act of faith, instead of science.

In the case of migration, available data sources for event recording are yet more limited. Vital registers by themselves do not register migratory moves, only full continuous population registration systems might do.

However, relatively few countries maintain such continuous population registration systems. Their design, implementation and maintenance are extremely costly. Furthermore, a minimum requirement for them to be effective as data sources is that the benefit of being and remaining properly registered is universally perceived by the population as outweighing the disadvantages. This is not always the case, and moreover the resulting non-registration and non-timely registration are often selective:

For example, recent, frequent, circulatory, illegal and undocumented migrants are typical cases in point where continuous population registration systems have a tendency to fail as adequate data sources on migratory events because of selectively incomplete or untimely registration.

Additionally, if the registration system employs a de iure concept of usual residence, then its value for the measurement of migratory events is also severely limited.

An example of such a de iure approach is the hukou system, the continuous household registration system maintained in China. Here, this has resulted in a major discrepancy between recorded (de iure) and actual (de facto) usual residency. As a consequence, for example tens of millions of migrants sustaining the current low-labour-cost economy of the country and their migration trajectories are effectively untraceable through the hukou system because of its de iure basis.

Further, for the continuous registration of migratory events, the traditional longitudinal household surveys tend to be a less useful (semi-)continuous substitute than they are for vital events and the associated vital event registers. There are two main reasons for this.

First, often these surveys are not nationwide. However, for a comprehensive insight into ongoing processes of migration, it is necessary to observe events that occur in all migration defining areas (MDAs).

Second, such surveys are commonly based on some form of random sampling.

Migratory events differ from vital events in that they are specific *both* by MDA of origin *and* by MDA of destination. Consequently, in order to reduce the uncertainty associated with random selection to acceptable levels, other things equal, sampling fractions required for the observation of migratory events have to be very much higher that those required for the observation of vital events.

There are two further issues of concern when counting the number of observed events for the traditional computation of empirical annual rates of the type M_x: the potential recurrence of events, and the incompleteness of the recording of events.

Unlike the event of dying, demographic events such as giving birth and making a migratory move are both *potential* and *potentially recurrent events*: They can happen 0, 1, 2, … times in the life history of cohort members. This causes methodological and operational complexities that are not similarly encountered in the case of mortality.

Methodologically, the issue revolves around the fact that the power of demography derives from the focus of its core paradigm: the study of demographic events as they occur in the life history of cohort members.

However, the mere counting during a year of the recorded number of events of birth or of migratory moves for the numerator of traditional rates of the type M_x does not result in an unambiguous relationship between the counted events on the one hand, and the life history of the cohort members that have actually experienced these events on the other.

This is, for instance, also illustrated by the fact that it is not uncommon in applied practice to obtain the event count for the numerator and the exposure for the denominator from entirely different sources. A typical example of such unrelated data sources is the use of a vital registration system for the event count, and the use of one or more population censuses for the assessment of the exposure. In such cases, the data sources employed simply prohibit associating the events that have occurred with the cohort members that have actually experienced the events in question at some point in time and age.

In fact, even if the event registration system employed as the source of data would actually allow one to obtain such information that links events and cohort members' life histories, then the mere computation itself of the observed annual fertility or migration rates of the type M_x breaks and discards those links:

The resulting annual fertility rates are average numbers of events per year per person at risk. They do not differentiate between women having given 0, 1 or multiple births during the year of observation. Nor do they represent any information on the actual point in age and time during which births, and any successive births, actually occurred to the cohort members involved.

Similarly, in the case of migration, the resulting annual migration rates themselves will have lost any information that might have been available from the data source used on the occurrence amongst cohort members of multiple moves and therefore also on the associated trajectories, on frequent moves, on their placement in age and time, and so on.

These are issues that are, of course, central to a full insight in and understanding of ongoing migration processes. Such deliberate loss of information is wasteful, and as we shall see later, it is also unnecessary.

The ability to count events for the numerator of traditional rates of the type M_x in the case of migration in such a manner that the relationship with the life histories of the cohort members is preserved, is further compounded by operational issues.

The main reason is that if a person experiences multiple moves during the year of observation, then these events may well be recorded in different offices of the registration system. Generally, of course, existing continuous civil registration systems have not been designed with the purpose foremost in mind to act as a data source for the measurement and analysis of migration.

In practice to this day, reconciling events with persons along age and time, and producing such data in an anonymized, cost-effective and timely manner and specific by migration defining areas specified on-demand, tends to be beyond the capabilities of most continuous population registration systems, however well automated.

However, in actual empirical application, perhaps the single most important practical issue regarding the counting of the number of occurrences of migratory events for the numerator of rates of the type M_x might well be *incompleteness*.

When counting the number of migratory events for the computation of such traditional annual migration rates, then a high proportion of the events that are to be counted will often concern recent migratory events. For a range of social, economic, cultural and legal reasons, varying from country to country and from person to person, recent migratory moves tend to be notoriously poorly recorded, whatever the data source used. Later, in our empirical example for Bangkok, Thailand, we shall for instance see that some 70% of all recent migratory events have in fact remained unrecorded.

Such proportions of incompleteness, of course, render any resulting traditional *migration rates* entirely *meaningless* empirically.

Moreover, and importantly, the traditional approach to evaluating observed demographic rates of the type M_x provides no warning against the potential occurrence of such incomplete counting of events. It neither provides any means to assess the severity of the problem, nor does it provide any methodologically justifiable manner in which to correct the resulting rates for such observation errors.

We shall return to this issue later. However, we note here that this same issue of incompleteness similarly affects migration data obtained from censuses or surveys when using a retrospective question on the place of usual residence a fixed number of years, usually one or five, prior to the enumeration. Here, too, the comparatively poor enumeration of recent migrants results in data that can be so seriously deficient as to make them empirically meaningless. And here too, the extent of the problem cannot be ascertained, nor are there any associated and methodologically sound procedures to correct for the resulting deficiencies.

Clearly, other things equal, the shorter the time interval chosen, then the more serious the issue will be. On these grounds alone, a question on the place of usual residence one year prior to the enumeration should normally be rejected. However, as we have seen earlier, this type of questioning suffers from significant other methodological drawbacks to the extent that this approach to measuring migration should in fact always be rejected.

The *occurrence* of events for the numerator is not the only issue that is problematic in the computation of traditional empirical annual demographic rates of the type M_x. There is also a major issue with the evaluation of the *denominator* of such rates: the ***exposure***, that is, *the number of person-years that cohort members are exposed to the risk of the demographic event in question*. We shall briefly explore this next.

The issue with the exposure is a problem that is commonly overlooked, and it is also methodological in nature. Specifically, the observed annual rates of the type M_x are not usually properly defined as occurrence / exposure rates. Let us again consider the familiar case of mortality. Even if the observed numbers of occurrences of events of dying are correctly allocated to the proper cohorts, and even if all of the above issues with the occurrence for the numerator would have been resolved, then there often remains a fundamental problem in the manner in which the exposure is measured.

Mathematically, of course, the exposure is an integral function of equation (6). In common life table notation, and consistent with standard practice in mathematics regarding the notation of primitive functions (integrals), this integral of ℓ_x is denoted by capital L_x. As we can see from equation (6), this exposure is in fact a function of the instantaneous rate $\mu(t)$.

In other words, *in order to be able to evaluate the empirical annual rate M_x, we first have to know the instantaneous rate $\mu(t)$* itself.

There is little doubt that this paradoxical situation is one reason why crude approximations of the exposure, such as the mid-year cohort size, remain universally popular.

In evaluating the set of M_x, one could, of course, assess the empirical exposure correctly: For instance, in the case of mortality, there is no difficulty in principle simply to sum the actual periods of time lived both by the survivors (one full year each) and by the deceased (parts of the year, varying from deceased to deceased depending on the date of death) during the year of observation. However, this is not standard practice.

In the case of official national statistics, one reason that this is rarely done in practice is not so much that the necessary information might not be available in the data sources from which official statistics are commonly derived. Often, in fact, it is. More usually, the key issue is that the sources in question are simply not designed to produce such data. As we have seen, to this day sources such as continuous vital and population registration systems tend to be designed as *registers* with restricted and clearly pre-defined administrative purposes in mind. However advanced the state of automation of such sources may be, they are not usually designed as modern *information systems* where timely and cost-effective flexibly-tailored on-demand data handling and information retrieval is a core consideration.

However, even if the empirical exposure would be correctly assessed as set out above, then at best one would still only obtain *stationary* (constant, static) *annual rates M_x* for each cohort. One would still not obtain the *instantaneous rates regimes $\mu(t)$* that are central to demographic theory. Such instantaneous rate functions $\mu(t)$ are not only conceptually and mathematically *proper cohort rates*, they are also *continuous functions of time and age*. So, clearly, by representing the *dynamics* of the underlying demographic processes over time and age, these functions $\mu(t)$ are also *much more informative* than any set of empirically observed static annual rates of the type M_x.

In summary, as we have seen, even in the conceptually simple case of mortality traditional methods of demographic measurement centring on empirically observed annual mortality rates are methodologically flawed: Often the resulting rates M_x used as the basic empirical input for further analysis are conceptually improperly defined, and never do they result in the full informative value inherent in demographic theory and in common data sets.

These problems are yet more acute when such methods of measurement are translated into the conceptually more complex areas of migration or fertility.

These are examples of demographic processes that are characterized by considerably *greater dynamism* from cohort to cohort and from time to time than is mortality. In addition, processes such as migration and fertility are demographically *more general* processes in that the events are potentially recurring during the life history of individual cohort members.

For a full insight into such processes, this potentially recurrent nature of the demographic events requires that both the data and the approaches to measurement *maintain the relationship* between the *events* on the one hand, and the *age and time life history of the cohort members* experiencing these events on the other.

The obvious approach to demographic measurement is instead to *abandon the use of empirical annual rates of the type* M_x altogether:

We recall once more that the essence of demography is the study of demographic events in the life history of cohorts. Logically, therefore, the key to demographic measurement is to chart exactly that.

Thus, instead of attempting to observe period rates as the point of departure, for individual cohort members one directly observes *the points in age and time at which events under study occur*.

Consider that we start observing a cohort at some given *benchmark point in time*. It is convenient to define this benchmark time point as $t = 0$. At this point in time, the cohort will be of some given age. For perfect cohort homogeneity with respect to age, we would of course require that all cohort members be of exactly the same age x. Clearly, however, actual empirical data sets rarely, if ever, allow for such perfect age homogeneity. In applied practice we therefore aim for individual cohorts whose members are closely similar in age at the benchmark point in time $t = 0$, that is, whose members at $t = 0$ fall in an age bracket $[x, x+a)$ where a is positive and as small as is reasonably and practically possible.

We then trace the life history of this cohort as time and age progress. The length of time during which we observe this life history is called the *observation time interval* or the *observation period*.

Specifically, for each cohort member, we then ***observe*** the *time interval that has elapsed*, that is, the *duration* from the benchmark time point $t = 0$, *until the first occurrence of the event under study* or *until the end of the observation period*, *whichever comes first*.

Let us give some examples. Consider, for instance, the case of mortality for a given birth cohort. Then this duration is the exact (that is, the not rounded) number of years lived by each cohort member until death or until the end of the observation time interval, whichever comes first. (It is useful to compare this with our discussion, above, of the measurement of the *exposure* in traditional M_x rates.)

In the case of migration and for a cohort, say, aged [20, 25), this duration is the length of time elapsed until the event (the migratory move) of interest or until the end of the period of observation, whichever comes first.

Depending on one's focus, the event of interest here may be the *first move* during the time interval of observation, but it may also be a *higher order move*, that is, a second, a third, …, move, during the observation time interval.

In fact, more generally, for the *complete observation* of *multiple events* (event recurrence, such as multiple moves in the case of migration) in the life histories of cohort members, we simply extend our observation *also to include* the *durations between successive events*, again until the end of our observation period. In the case of migrations, such durations are commonly known as *residence durations*.

Methodologically, such complete data sets open up the possibility of a further range of in-depth approaches to measurement. Key approaches are detailed in Xu-Doeve (2006). In our subsequent discussion here, however, we shall limit ourselves to first events (in the case of migration: first moves) only.

Next, the resulting duration data t*ranslate directly into the cohort size unaffected by the event in question as time and age progress*. In other words, we can now directly obtain a *sequence of monotone decreasing data points* representing the *number of cohort members that have not (or not yet) experienced the demographic event* under study as time and age progress:

Formulated generally, at the start of our observation time interval we begin with a cohort of size $K(0) = K_0(0)$. Recall here that $K_0(t)$ denotes the number of cohort members at time point t who have experienced zero events on the time interval running from the benchmark point in time (point in time zero) until time point t.

Then at each point in time t at which the event in question occurs to a cohort member we, of course, have that the number of cohort members having experienced zero events is reduced by one. Thus, our empirical observation of durations until the occurrence of an event results in a sequence of data points $(t, K_0(t))$ over the time interval of observation. *This set of data points* $(t, K_0(t))$ *is the empirical expression of equation (6)*, above.

In the familiar case of mortality, these data points represent the empirically observed number of remaining survivors as age and time progress. Expressed more formally, they are *alive-state (or, more generally, origin-state) attrition data*. Thus, for a cohort of perfectly homogeneous age such a data set represents proper actual *empirical cohort data* ℓ_t, rather than synthetic cohort data. Note that, if we start observing this cohort at exact age zero, then $\ell_t = \ell_x$.

In the case of migration (and in most other cases, such as fertility), the cohort under study will also have been subject to exposure to the *competing risk* arising from the *simultaneously operating force of mortality* over its observed life history.

Depending on the data source used -- and, such as in the case of a continuous population registration system, also depending on the precise manner in which data are actually retrieved from this source --, the effect of this competing risk of mortality may or may not be represented in the resulting set of $(t, K_0(t))$ data.

If indeed it is represented in this data set, then the monotone decline exhibited by the data points as time and age progress is a *compound effect of the simultaneous operation of the forces of mortality and out- or emigration*. When we are concerned with the measurement of migration, then, of course, we wish to obtain *data that reflect exclusively* this *process of migration*.

Consequently, in such cases the observed origin-state attrition data must be adjusted for the effect of this competing risk of mortality. Such elimination of the effect on the data of the force of mortality is commonly achieved by a simple elementary standard *reverse survival* procedure, using available information on the mortality schedule of the cohort under study. More details are available in Xu-Doeve (2006).

Let us recall that the resulting sequence of data points $(t, K_0(t))$ -- after adjustment for the effect of the competing force of mortality where appropriate -- is *the empirical manifestation of* $K_0(t)$ *in equation (6)* of demography's mathematical theory.

Note that the demographic behaviour of the cohort as embodied in equation (6) is determined *exclusively* by the demographic force $\mu(t)$ that governs the process in question, or, equivalently, by the instantaneous rate $\mu(t)$. Therefore, we now have the empirical information that enables us to quantify this instantaneous migration schedule $\mu(t)$. However, here we are faced with what initially might seem a difficulty.

As we have seen, *demographic theory leaves the functional specification of instantaneous schedules* $\mu(t)$ *undefined*: It does not prescribe one or more particular classes of functions, except that they be continuous, real-valued and non-negative.

Also, *empirical research* over the years has never convincingly succeeded in producing any *generally-valid* functional specifications for mortality, fertility or migration schedules which adequately capture the lifetime mortality, fertility or migration experience exhibited by human cohorts.

In such cases, a common standard approach in science is *piecewise polynomial approximation*: We abandon trying to specify some single lifetime mathematical form for $\mu(t)$. Instead, we *partition* the interval of observation into a set of subintervals such that a *comparatively simple polynomial function adequately captures the demographic behaviour locally on each subinterval*.

In empirical practice in the case of migration, the following rule of thumb is generally satisfactory:

1 the *subintervals* must *not exceed 5 years in length*

and this is combined with

2 a *quadratic, cubic or quartic* local polynomial specification of $\mu(t)$.

Note that the local polynomial specification that performs best *may vary from subinterval to subinterval for a given cohort*, and that the best performing set of local polynomial specifications *may vary from cohort to cohort*. For each individual subinterval, there are two factors to consider here.

The first criterion is, of course, *goodness of fit*: How well does the chosen specification of $\mu(t)$ reproduce the observed data on a given subinterval. The second issue to consider is, whether or not it is desirable to allow for some measure of *smoothing* of the observed data to account for known errors in the observed data points $(t, K_0(t))$. One example of a case where smoothing might be indicated is where it is known that there exists digit preference in the reported duration data. Note, however, that smoothing by its very nature implies a less than perfect fit. These two criteria must therefore be assessed in a well-considered and balanced fashion.

On many age ranges, quadratic and sometimes even linear polynomials may already give reasonable to good results. In fact, because of their increasingly strong oscillating tendency, polynomials of a higher degree than quartic are not normally recommended. If the fit of all polynomials that are specified in the rule of thumb, above, is less than adequate, then the choice of *shorter subintervals* than 5 years is indicated instead.

Let us briefly explore the operational procedure underlying the general method of demographic measurement in some more detail in formal terms; then, in the next section we shall illustrate matters using empirical data.

Consider that we, for example, select a cubic polynomial specification of $\mu(t)$ on a given piecewise observation interval. Then we have

$$\mu(t) = \beta_1 + \beta_2 t + \beta_3 t^2 + \beta_4 t^3 . \quad (7)$$

Integrating equation (7) over time interval [0, t), we obtain:

$$\int \mu(u)du = \beta_1 t + \tfrac{1}{2}\beta_2 t^2 + \tfrac{1}{3}\beta_3 t^3 + \tfrac{1}{4}\beta_4 t^4 . \quad (8)$$

Next, substituting equation (8) in equation (6) gives:

$$K_0(t) = K(0)\exp\{-(\beta_1 t + \tfrac{1}{2}\beta_2 t^2 + \tfrac{1}{3}\beta_3 t^3 + \tfrac{1}{4}\beta_4 t^4)\} . \quad (9)$$

We recall here that $K(0) = K_0(0)$. Let us now write $\exp(\beta_0)$ for $K(0)$, and let us write $\exp\{y(t)\}$ for $K_0(t)$. Then, taking logarithms, we finally have

$$y(t) = \beta_0 - \beta_1 t - \tfrac{1}{2}\beta_2 t^2 - \tfrac{1}{3}\beta_3 t^3 - \tfrac{1}{4}\beta_4 t^4 , \quad (10a)$$

or, more clearly identifying the independent variables of the RHS in brackets, and therefore more conveniently

$$y(t) = \beta_0 + \beta_1(-t) + \beta_2(-\tfrac{1}{2}t^2) + \beta_3(-\tfrac{1}{3}t^3) + \beta_4(-\tfrac{1}{4}t^4) . \quad (10b)$$

Next, we also take logarithms of our empirically observed origin-state attrition data $K_0(t)$. Recall that these data items represent the observed *number of cohort members who have not (or not yet) experienced the demographic event under consideration*, expressed as a function of time t. Thus, in the case of migration, these $K_0(t)$ values are the empirically observed *numbers of cohort members not (or not yet) having left* (or, more informally, still remaining in) *the MDA of origin as age and time progress*.

Then the issue of mathematically and methodologically sound and maximally informative *demographic measurement* simply *reduces to estimating the coefficients of equation (10b)*, *given* the empirical log-data, that is, given *the logarithms of the observed $K_0(t)$ values*. Statisticians often refer to this procedure loosely as "curve fitting".

In the case of a complete enumeration of the cohort, one can use *ordinary least-squares estimation* well-known from text-book multiple regression analysis. When, on the other hand, the empirical data have been obtained using random sampling, then *maximum likelihood estimation* is the recommended procedure. Such estimation procedures can be performed as a matter of routine using any well-validated statistical software application.

Next, the resulting values for coefficients β_1, β_2, … can now be substituted in equation (7) which specifies the demographic force function $\mu(t)$ or, equivalently, the instantaneous rate function $\mu(t)$ on the piecewise observation interval. In the case of migration, this gives us the desired *instantaneous migration rate $\mu(t)$ as a continuous function of time and age*.

Further, recall that $K_0(0) = K(0) = \exp(\beta_0)$. Also, recall that the instantaneous rate functions $\mu(t)$ are defined specific by MDA-to-MDA trajectory. Therefore, we can now also obtain *absolute numbers of migrants specific by both cohort and trajectory* (absolute flows and stocks) *as a continuous function of time and age* on any time interval $[t_1, t_2)$: Using equation (9), we can directly evaluate the difference $K_0(t_1) - K_0(t_2)$.

As discussed, if our originally observed data reflect the competing forces of both mortality and migration, then as a first step in our measurement procedure we eliminate the effect of mortality on our data set. Consequently, the resulting absolute flows and stocks are *before any mortality*.

An interesting case arises if we select a polynomial of the lowest degree (that is, of degree zero, and thus a constant) as the local specification for $\mu(t)$. In other words, let us suppose that we select $\mu(t) = \mu = \beta_1$. Then, of course, the result of the measurement procedure on the subinterval will be a single average best fit *point value* for the instantaneous rate. Clearly, such a point value cannot and does not represent any of the actual *dynamics* of the demographic process on that subinterval *as time and age progress*.

From the point of view of the resulting informative value, such a piecewise constant rate on any given time and age range is quite similar to a traditionally measured empirical annual M_x type of rate. However, unlike such a traditional empirical rate of the type M_x, this resulting value of μ will be conceptually a proper cohort rate and it will be mathematically consistent and methodologically sound.

It will be clear that demographic measurement centres on the observation of demographic events experienced by cohort members as time and age progress. More specifically, the accurate recording of the *timing of events* as these occur in the individual life histories of cohort members is the basis of the measurement of demographic processes.

In the case of vital event registration, the *time resolution* (the level of detail on the time scale, and thus implicitly also on the age scale) in many countries is very high: not uncommonly, the events of death and birth are recorded with a detail down to hours and minutes on the day. However, such a high level of time resolution is not practically available for the recording of migratory events.

In addition, very few countries have continuous population registration systems that allow the retrieval of timed migratory events. Consequently, the usual national source of migration data will be the periodic population census.

In order to obtain information on migratory events in the life history of persons in a census, it is necessary to employ retrospective questioning on residence durations or, equivalently, on dates of migratory events in the past. The ability to recall the exact timing of such historical events is, of course, limited.

Yet, it will be clear that the greater the accuracy and detail of the answers, then both the more precise and the more informative the resulting insights into ongoing migration processes will be. Intelligent probing by expert census field professionals greatly facilitates such accuracy and detail.

In the increasingly common case of self-administered census questionnaires, however, control over response quality is very much left to the respondent. From the point of view of the measurement of migration, this is a matter of concern.

In the case of retrospective questioning, when using the standard 5-year wide piecewise observation intervals, then the absolute *minimum requirement* as to time resolution is *migratory event dating precise in single years*.

If the time resolution used for event observation is cruder than this absolute minimum of single year precision, then insufficient numbers of distinct data points (t, $K(t)$) will result, effectively *denying a meaningful measurement procedure*.

A precision down to the level of *quarters* or *seasons* is already a significant improvement. However, truly and generally satisfactory for the measurement and analysis of migration is precision down to the level of *months*, and this should be the *norm in literate and largely formalized societies*.

A further issue of special significance in the case of the measurement of migration is the matter of the incompleteness of observation. Specifically, it is common that *empirical data on recent migrants and their migratory behaviour* are of a *comparatively poor to very poor degree of completeness*.

There are several reasons -- such as administrative, socio-economic, cultural, and legal -- why recent migrants are disproportionally difficult to observe empirically. We shall not discuss these factors here. However, we do note that the longer a migrant remains in his or her destination place of residence, then the less the person in question will differ from a longer-term resident, at least from the point of view of the likelihood of being captured in a statistical data collection process.

And thus, for example in a population census where migratory events are recorded by questioning after the event (retrospective questioning), *the difference in the* *"capture-rate"* between *migrants* and *non-migrants* within an MDA *will vanish with increasing length (duration) of residence*.

This is an issue that is of direct relevance in the measurement procedure: This disproportionate degree of incompleteness of recent arrivals is often so serious that in the above procedure to estimate the coefficients β_n of the instantaneous rate function $\mu(t)$ *data points (t, K(t)) pertaining to the most recently arrived migrants have to be discarded*.

Once the coefficients have been estimated without using these discarded data, and again recalling that $K(0) = \exp(\beta_0)$, then the *incompletely observed data* on absolute numbers of recent arrivals can be *adjusted* using equation (9) above. We shall see a worked example of this adjustment procedure in the empirical analysis presented in next section.

The use of *population census data* for the measurement of migration processes is in fact special in a very particular manner. As will be clear from the discussion above, the demographic measurement procedure as outlined centres on the occurrence of events that *reduce* the number of cohort members who have experienced zero events as time and age progress. In the case of migration, these are events of *outmigration* or *emigration*.

However, when we use retrospective questioning in a population census on the duration of residence in the present place or country of usual residence, then we are effectively observing events of *inmigration* or *immigration*, instead.

Methodologically, the procedure to deal with this is <u>*time reversal*</u>, and we shall discuss this too in the next section where we shall present an empirical example of the measurement method using population census data from Thailand.

5 Applied Measurement Using Population Census Data

As we have seen earlier, for most countries, not only in the developing world, the periodic population and housing census is the prime -- if not the only -- source for a comprehensive national data set on migration. In this section, we shall therefore illustrate the methods developed in the previous sections using empirical data derived from a population census.

An understanding of how to employ a population census for the measurement of migration is especially important at the present time: At the time of writing, preparations for the *2010 global round of population censuses* are well under way, and several countries have in fact already taken their "2010" census.

Unfortunately, there are many instances worldwide of past censuses that have proved less than optimal from the point of view of measuring migration. Inevitably, suboptimal conceptual, methodological and / or operational practice has long-lasting consequences:

Most countries take a population census only once every ten years. Therefore, if the approach adopted to the measurement of migration in the present census is conceptually, methodologically or operationally flawed, then this may effectively mean that a country will have to wait until the 2020 census round when it will first be able to collect a new data set that does allow meaningful and informative measurement and analysis of ongoing migration processes.

Thus it is important that the key issues of *concepts and operational definitions*, of *methods of measurement*, and of *operational practice* in *data collection* and *information systems design* are given careful consideration in the preparations for the present census round.

We recall that the *measurement of migration* fundamentally centres on establishing the *timing*, *origin* and *destination* of ***migratory events*** (***moves***) *in the individual life histories of cohort members*.

As mentioned, in a population census, such information is obtained by *retrospective questioning*. An associated benefit of retrospective questioning is that life history data of respondents are obtained direct, so that migration event histories along cohort lines can be assembled unambiguously.

The standard census question allowing the recording of the required information for the most recent migratory event is the question on the ***duration of residence in the present place*** (or, for international migration: ***country***) ***of usual residence***.

It is important to note that *in the absence of this (or an equivalent) question* that establishes the *timing of the migratory events under study in the observed life histories of respondents*, it is *impossible properly to measure migration processes*.

Further, in order to establish the trajectory involved in the move under study, a companion question should be asked on the ***associated previous place*** (or, for international migration: ***country***) ***of usual residence***.

Such questioning can be extended to prior migratory events. For a more informative insight into ongoing migration processes a *key recommendation* is that *questioning be extended to capture* **at least the two most recent events *(moves)***, rather than merely the single most recent one.

Capturing more than just the last move is essential for exploring crucial issues such as *temporary migration*, *short-term migration*, and *frequent migration*. It is equally essential for obtaining in-depth insights into migration trajectories, including for example those involved in *circular and return migration*.

Clearly, extending the questioning in censuses to capture more migratory events than just the last move increases the census burden, both in terms of the number of questions to be asked and in terms of the necessary professionalism and expertise of the field staff. However, this can often easily be accommodated for by a fundamental review of core census objectives and procedures:

From the point of view of information gathering, many modern population and housing censuses are abused to a greater or lesser extent as an instrument of data collection. As discussed, in the absence of an adequate system of population registration, a *full population census* is inevitably necessary for the establishment of a comprehensive insight into ongoing processes of migration that affect and shape countries. However, in this respect, migration is, in fact, fairly exceptional:

Adequate and reliable information covering many of the topics commonly included in modern population and housing censuses can be obtained more efficiently, more cost effectively and more timely by using *sample surveys* and / or from *other sources*. Obvious examples of such other sources include, for instance, existing administrative data, remote sensing and relevant service providers.

Cleaning up census questionnaires by removing such items and by centring exclusively on those topics for which a full population enumeration is indispensable, can free up the census programme significantly. This allows for the accommodation of more in-depth questioning on issues such as migration. At the same time it facilitates the development of a more focused professional expertise among field staff.

As far as the measurement of migration is concerned, methodologically deficient and informationally poor measurement instruments in population censuses *should be abandoned* in favour of the above *event-centred questioning*.

Such deficient and poor instruments include, for example, *person-centred questions*, such as questions on the *place or country of usual residence at some fixed point in time in the past*, questions on *place or country of birth*, and questions on *nationality or citizenship*.

Earlier, we already discussed in some detail the fundamental limitations and drawbacks of questioning on the place or country of residence at a fixed date in the past. We also note that maintaining such a question simply in order to ensure continuity with earlier censuses is a spurious argument. Given observed instantaneous migration rate functions $\mu(t)$ and using demography's mathematical theory, the distribution of cohorts by place of residence not just at one-year or five-year intervals but at any point in time t can be directly obtained.

A question on the place or country of birth equally does not result in the observation of any migratory events in the life history of persons. For the measurement of migration, such questions are therefore clearly unsuitable.

Questioning on nationality or citizenship in order to measure international migration is demographically flawed as well, since here too no events are actually observed. This, alone, is enough to render it impossible to derive dependable information on ongoing processes of international migration. It might instead perhaps be argued that such data do have sociological and political significance.

However, a close examination of the definitions of the concepts actually involved, and of the unstable nature of these concepts of nationality and citizenship from country to country as well as over time and in the life history of individual persons, will immediately reveal that the resulting data are *at best* difficult to interpret unambiguously. Consequently, this leaves such data open to unverifiable misinterpretation and misrepresentation (Xu-Doeve, 2006).

When it comes to the measurement of migration, the selection of appropriate questions to be included in the census is not the only issue to consider. Probably more than any other topic included in a population census does migration place demands on the *information system* used for data storage, organization and retrieval.

The delivery of timely and relevant information on ongoing migration processes for use in policy making, planning, monitoring and evaluation *entails significantly more than* the *production of a set of the well-known routine traditional standard census tabulations*.

It requires the ability *flexibly* and *dependably* to assemble cohorts that are specified *on demand*, and whose members' timed migratory events are specific by migration defining areas (MDAs) that are defined *tailor made* to the particular issue under investigation. It may require the assembly of such cohorts at points in time other than at $t = 0$. If more than just the most recent move has been observed, then the generation of appropriate data sets is yet more demanding. Explanatory studies exploring relationships with other covariates further complicates the demands placed on information management and retrieval. And, for example, census linking, enabling the more reliable compilation of longer life history event time series, poses additional challenges to information system design and management.

National statistical offices (NSOs) must have the capacity to perform tasks such as the production of the necessary data sets *on request* and *as a matter of routine*, while *dependably meeting the required specifications* of the data set, and in a manner that is *timely*, *efficient* and *cost-effective*.

It might seem attractive to opt for one of the established low-cost or free census data entering, editing and tabulating applications, such as CSPro, promoted by the United States Census Bureau, or the older IMPS and ISSA packages. However, here, the sting is in the long post-acquisition tail.

First, these applications are not designed with the requirements of modern applied demographic measurement and analysis in mind. Consequently, meeting even the most basic of the tasks outlined above proves highly cumbersome, time consuming and error prone.

Second, the approach to data management of these applications does not follow standard principles of modern information systems design common throughout the information technology sector. As a result, they require a high degree of application-specific expertise and skills.

Overall in the final analysis, therefore, these applications may well be low-cost up-front, but they are both *poorly equipped for the task* and *ultimately costly to operate*. Xu-Doeve (2006) discusses the question of appropriate information systems in detail and suggests alternative and future-proof solutions.

Before we can proceed to the presentation of our empirical case, there is one more methodological issue that needs our attention. We already alluded to this at the close of the previous section. It concerns the matter of *retrospective data on migratory events* in the life histories of cohort members typically encountered in data sets derived from population censuses.

Retrospective census data on residence durations are typical in that such durations are observed for *in-migrants and immigrants* into the MDA of enumeration. As we have seen, demographic theory is formulated instead in terms of attrition of cohort numbers resident in an MDA as a consequence of *outmigration and emigration*. This paradox is easily resolved, however, by the device of *time reversal*.

Imagine, for instance, that the process of "border" crossing (the migratory events) *by arrivals from one MDA into another* had been filmed in real time. Then the resulting motion picture would show the *increments*, as time and age progressed (increased) up to the point in time of the census enumeration, to the numbers of cohort members resident in the MDA of destination as a consequence of in-migration or immigration.

Now, if we next play this film backwards, that is, if we *reverse the order of time*, then we actually see *the very same process* but this time as *decrements* gradually depleting the number of cohort members resident in the MDA of destination as time and age progress backwards (decrease) from the time point of the census enumeration.

In other words, by such time reversal, the migratory events that have *actually* been *recorded as events of arrival* are *formally represented as events of departure*.

Thus, for example, by reversing the order of time we obtain instantaneous migration rates $\mu(t)$ that actually are in- or immigration rates, but which have merely been measured technically as if they were out- or emigration rates.

It is important to realize that, as a consequence of this device of time reversal, a time point, say, $t = 3$ in any equation refers to a point in time 3 years *prior to the enumeration*.

This concludes our overview of conceptual, methodological and operational issues that are of special concern and of specific relevance when we measure migration using population censuses.

Next, we shall briefly present selected empirical results using data from a population census, namely the 1970 Thai Population and Housing Census.

The data are for migration into Bangkok from the rest of the Kingdom for the male cohort aged [20, 25) at the time of the enumeration (t = 0). Observed residence durations are classified in 0.5-year wide categories[1].

Such data are available up to 5 years prior to t = 0. Consequently, recalling our above rule of thumb for the piecewise approximation of $\mu(t)$, in this case we only have one single 5-year wide piecewise interval of observation.

The following table and graph present both the observed data and the measurement results.

Table 1 The Measurement of Inmigration into Bangkok:
Male Cohort Aged [20, 25) at the Time of the Census
(Cohort Sizes and Migrant Numbers × 100)

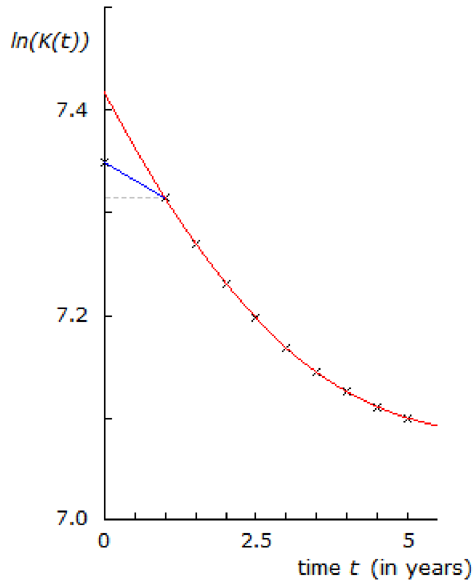
Time t	Observed Migrants on $[t, t+0.5)$ *	$K_0(t)$ **	$ln(K_0(t))$	Estimated $ln(K_0(t))$	Estimated $K_0(t)$	Estimated Migrants on $[t, t+0.5)$ ***
0	52.1	1553.4	7.34820	7.41746	1664.8	87.8
0.5				7.36329	1577.0	75.8
1	65.0	1501.3	7.31409	7.31406	1501.3	64.8
1.5	54.5	1436.3	7.26983	7.26994	1436.5	54.9
2	46.0	1381.8	7.23114	7.23100	1381.6	45.8
2.5	38.0	1335.8	7.19729	7.19726	1335.8	37.7
3	30.1	1297.8	7.16843	7.16861	1298.0	30.4
3.5	23.8	1267.7	7.14496	7.14490	1267.6	23.9
4	18.5	1243.9	7.12601	7.12587	1243.7	18.2
4.5	13.0	1225.4	7.11102	7.11115	1225.6	13.2
5		1212.4	7.10036	7.10033	1212.4	

$$\mu(t) = 0.11310 - 0.01872\,t - 0.00121\,t^2 + 0.00023\,t^3 \qquad (R^2 = 0.999998)$$

* After adjustment for mortality; 52.1 refers to time interval [0, 1)
** $K_0(t+a) = K_0(t) -$ Migrants on $[t, t+a)$, given observed $K_0(0) = 1553.4$
*** Migrants on $[t, t+a) = K_0(t) - K_0(t+a)$
Note: Time $t = w$ refers to w years *prior to* the enumeration

1 Please refer to the appendix at the end of this section for more details on the observed duration of residence data presented here.

FIGURE 1 The Measurement of Inmigration into Bangkok:
Male Cohort Aged [20, 25) at the Time of the Census
Observed Data Points (t, $ln(K_0(t))$) and Best-Fit Curve for Cubic $\mu(t)$



Because we have a full enumeration (no random sampling), we used least-squares estimation to obtain the coefficients of equations (10b) and (7).

It will be clear from both the table and the graph that the quality of event recording on the residence duration interval [0, 1) has been highly defective. As discussed, we therefore *omitted the data for this interval* in the least-squares coefficient estimation procedure: We used only the nine data points (t, $K_0(t)$), or, after taking logarithms, the nine data points (t, $ln(K_0(t))$), that are available for the residence duration interval [1, 5).

The resulting empirically measured function $\mu(t)$ specified as a cubic polynomial proves to give an excellent representation (R^2 = 0.999998, or 99.9998%) of the underlying *force of migration*, and consequently of the equivalent *instantaneous inmigration rates* of the cohort under study as time and age progressed.

This excellent fit also results in a very close agreement between the observed and expected absolute numbers of cohort members by duration of residence (*migrant flow and stock data*) over the interval of observation.

The important exception here is the recent [0, 1) residence duration category, where we see a major discrepancy between the observed and expected numbers of migrants. This is, of course, not surprising given the poor quality of the empirical data for this category comprising the most recent arrivals:

We recall our earlier discussion of this phenomenon: Here we see high levels of *migration-specific incompleteness*, in the enumerated data, that is, incompleteness related to the recentness of the event of inmigration.

Specifically, according to the measurement results obtained above for this group, $\{(87.8 + 75.8) - 52.1\} / (87.8 + 75.8) = 0.68154$, or *nearly 70% , of these recent migrants have been missed in the census*.

We can, however, take this analysis of the underenumeration of migrants one step further. Any properly conducted census is subject to a quality assurance process, and a quality assessment was also carried out for the 1970 Thai Population and Housing Census. It was based on a post-enumeration survey, supplemented with additional quality and consistency checks.

One result was, that the observed number of members of our male cohort aged [20, 25) at the time of the enumeration had been significantly underenumerated. Specifically, it was found that the observed number had to be adjusted upwards by a factor of 1.2819. This corresponds to an underenumeration rate of 21.99%. This result we use in the following table.

TABLE 2 The Measurement of Inmigration into Bangkok:
Full Adjustment for Underenumeration
Male Cohort Aged [20, 25) at the Time of the Census
(Cohort Sizes and Migrant Numbers × 100)

DOR	[0, 0.5)	[0.5, 1)	[1, 1.5)	[1.5, 2)	[2, 2.5)	[2.5, 3)	[3, 3.5)	[3.5, 4)	[4, 4.5)	[4.5, 5)	[5, ∞)	Total
Obs	52.1		65.0	54.5	46.0	38.0	30.1	23.8	18.5	13.0	1212.4	1553.4
Est I	87.8	75.8	64.8	54.9	45.8	37.7	30.4	23.9	18.2	13.2	1212.4	1664.8
Est II	105.0	90.7	77.5	65.7	54.8	45.1	36.4	28.6	21.8	15.8	1450.2	1991.3

The header row of table 2 gives the duration of residence ("DOR") categories. The first row of data ("Obs") are the observed data from table 1, and the second data row ("Est I") are the estimated data obtained in the first step of our estimation procedure, also from table 1.

Recalling that the cohort size numbers in the tables have to be multiplied by 100, the final measurement results of the migrant flow and stock data ("Est II" in table 2) are obtained in a second step in which we proceed as follows:

First, the observed total number of cohort members 155340 is adjusted using the above multiplication factor of 1.2819. This results in a total cohort size of 199130. Thus, in the census 199130 − 155340 = 43790 persons in this cohort have not been observed.

However, in step 1 of our measurement procedure we have already recovered 11140 (namely, 166480 − 155340) of these 43790 missed cohort members.

Now, recall our discussion earlier about migration-specific underenumeration as underenumeration directly related to the recentness of migratory events. Then, apart from the minor differences obtained in step 1 between observed and estimated numbers for the duration of residence categories from [1, 1.5) to [5, ∞), this recovered number of 11140 cohort members is accounted for by the *adjustment for migration-specific underenumeration in step 1* affecting duration of residence category [0, 1) representing recent migrants.

This leaves us with a remaining deficit of 199130 − 166480 = 32650 cohort members *not enumerated due to all causes of underenumeration other than migration-specific causes*.

By implication, there are no grounds to allocate these 32650 persons to any specific duration of residence category in preference over other such categories. Therefore, we distribute them proportionally over all duration of residence categories by applying a step 2 multiplication factor of 199130 / 166480 = 1.1961 to the data obtained in step 1 for all categories. The results of this final adjustment are shown as the bottom row ("Est II") of table 2.

It is interesting now to revisit the data on those recent migrants who have arrived within the 12 months immediately prior to the taking of the census. In the enumeration, 5210 of them were actually counted. However, table 2 shows that the true number was instead 10500 + 9070 = 19570. Thus, after full adjustment for underenumeration, we find an underenumeration rate for recent migrants of 73.38%.

In other words, for our cohort, usually one of the most mobile cohorts in a population, we find that in fact *nearly three-quarters of all recent migrants* have actually been *missed in the census enumeration*.

Recall, here, that the underenumeration rate for the cohort as a whole was 21.99%. The major difference between these rates illustrates the *disproportional* *propensity of recent migrants to be incompletely enumerated.*

We note that, although the underenumeration rates vary from cohort to cohort, similar major disproportional underenumeration of recent migrants is measured in all male cohorts in this population (Xu-Doeve, 2006).

More generally, this serves as a reminder that *taking observed data on migrants as valid and true can lead to a serious misrepresentation and misinterpretation of actual empirical reality*.

Furthermore, we reiterate that *measurement results such as those presented here* can be obtained *only if migratory events (migratory moves) are recorded as they occur in the life histories of cohort members as time and age progress*.

In a population census, such data are recorded *only if either residence durations or timed migratory events are observed* through appropriate retrospective questioning.

We conclude our discussion of the applied measurement of migration with a note of caution. The basic principles of demographic measurement are quite elementary and straightforward.

However, the path to applied practice, from conceptual and operational definitions through actual data collection to information processing, is a road with many -- and not always obvious -- pitfalls and obstacles.

Errors along the way can all too easily have serious consequences which in hindsight cannot anymore be rectified. The end result of the costly data collection process that a population census inevitably is, may then well be that a meaningful measurement and analysis of ongoing migration processes actually turns out to be seriously impaired or even impossible.

For national statistical offices that have limited or no experience with the modern measurement of migration, it is recommended that the analysis such as the one presented above is replicated, based on and starting with an *unprocessed full census* in which duration of residence data have been collected and for which all standard documentation, such as data definitions, enumerator instructions and code books, is available.

Critically comparing and contrasting the experience of such an exercise with own past practice and with national information requirements for the future is of immense value and can help avoiding bottlenecks and pitfalls in the concepts, questions, methods, processes and practices adopted for the own next national enumeration.

<u>Appendix</u>: Note on the Observed Duration of Residence Data

Residence duration data specific by one-year wide time intervals (and thus, implicitly, the derived $K_0(t)$ data at the annual time points) were empirically observed in the census.

Each of these observed annual categories, except the first, was subsequently broken down into two semi-annual categories, using 3rd-degree polynomial interpolation on the cumulated data, as part of wider research into the measurement of urbanization in Thailand at the request of Thailand's National Economic and Social Development Board in 1982. The data disaggregated by these semi-annual residence duration categories are presented in table 1 under the table heading "Observed Migrants on [t, t+0.5)".

However, only the underlying original residence duration data by annual categories [0, 1), [1, 2), …, [4, 5) constitute official Thai government statistics.

Note that such prior interpolation is not necessarily recommended practice. It is instead recommended directly to observe durations of residence themselves with greater time resolution in the census data collection process.

While linear and quadratic specifications of $\mu(t)$ for the cohort under study are given in Xu-Doeve (2006), the cubic specification is not, because of the limited number of available official observed data points (t, $K_0(t)$). In order nevertheless to be able to demonstrate the use of this important cubic specification of $\mu(t)$ without approaching a saturated coefficient estimation scenario, we present the interpolated data here.

CONCLUSIONS

To conclude this presentation of the basic principles of the measurement of migration, we shall briefly itemize -- without further discussion -- some of the principal results that are key to the successful measurement of migration processes.

In these conclusions we shall focus on the measurement of migration using population censuses as the source of the empirical data. Most of the conclusions, however, readily generalize to other data sources, as well.

The elementary and primary objective of the measurement of migration processes is to obtain *valid and reliable information*

1 on the *intensities* of these processes, that is, on *instantaneous migration rates*,
2 on the *flows and stocks of migrants in absolute numbers*, and
3 on the *dynamics over time and age* in these intensities, flows and stocks,

specific by the *individual migration trajectories* (geographical paths) under study. More in-depth measurement, as well as analysis of the findings, build on this foundation.

Such measurement of migration processes fundamentally starts from and centres on establishing the *timing*, *origin* and *destination* of migratory events (migratory moves) *in the individual life histories of cohort members*.

Census questions that do not measure this information are unsuitable for the measurement of migration. Such unsuitable questions include, for instance, questions on the place of usual residence at some fixed date in the past, questions on the place of birth, and questions on nationality or citizenship. None of these questions captures the timing and direction of any migratory events (moves) experienced by the persons involved. They merely record information that may in one way or another be related to such moves.

The standard census question allowing the recording of the required information on migratory event timing for the most recent migratory event is the question on the *duration of residence in the present place* (or, for international migration: *country*) *of usual residence*.

In order to establish the trajectory involved in the move under study, a companion question should be asked on the *associated previous place* (or, for international migration: *country*) *of usual residence*.

Such questioning can be extended to prior migratory events. For a more informative insight into ongoing migration processes a key recommendation is that questioning be indeed extended to capture *at least* the *two most recent migratory events* (moves), rather than merely the single most recent one.

Further, it is essential that actual *true durations of residence* be observed with as much precision as possible during the enumeration process through the use of appropriate measurement instruments and techniques. For example, c*ross-checking* residence *durations* with the *dates* of the associated migratory moves is a valuable device.

For the observation of the timing of events (residence durations, dates of migratory moves) a *measurement scale* which is precise down to *month and year* of arrival is satisfactory.

As to the issue of what in fact constitutes *usual* residence, the recommendation is to consider *every de facto change of residence* as a change of usual residence, *unless* it is *both short*, that is, *under one month*, *and at the same time* for a *purpose of stay that clearly suggests that pre-existing usual residence has not changed*.

The recommended *exhaustive list of purposes of stay* that *disqualify* such short residence as usual residence, comprises: recreation, tourism and holidays; business travel; temporary visits to friends and relatives; temporary medical treatment; and religious pilgrimage.

It is important that places and countries of usual residence be recorded with *sufficient geographical detail*, prepared for future information demands and anticipating the ability flexibly to compile tailor-made migration defining areas specified on demand by data users.

The measurement of migration places comparatively high demands on the information system used. The issue of appropriate census *database design* should therefore be given careful consideration. The use of *industry-standard database management systems* for data organization, entry, management, back-up, and retrieval is recommended.

These recommendations may well appear to add to the overall *census burden* in a country. Importantly, however, in the absence of an adequate system of population registration, a *full population census* is *inevitably necessary* for the establishment of a *comprehensive insight into ongoing processes of migration* that affect and shape countries.

Such a full enumeration is not necessary, however, for many of the topics commonly included in modern population and housing censuses.

Adequate and reliable information covering *many topics* -- including also many of the so-called United Nations core topics -- can be obtained more efficiently, more cost effectively and more timely by using *sample surveys* and / or by using *data from other sources*, including, for instance, existing administrative data, remote sensing, and relevant service providers.

Removing such topics from the census programme generates the necessary *space for the collection of more in-depth data on migration*. At the same time, it generates the space and opportunity among census field staff further to *develop the professional expertise and skills that are necessary for high-quality migration data collection*.

In 1970 the United Nation published its highly influential Manual VI on the measurement of migration: United Nations (1970). This manual focused primarily on a largely qualitative appraisal of direct questions and of sources, on various descriptive statistical indicators such as crude rates, ratios and indices, and on a thorough presentation of methods of indirect estimation of net migration. It was also limited exclusively to internal migration. As regards the direct measurement of migration, Manual VI lacked a rigorous and coherent theoretical and methodological framework.

Since that time, the field of the measurement of migration has witnessed radical development. It is a development that may be characterized as a shift from *measurement-based analysis and theory development* to *theory-based measurement and analysis*.

In the previous sections we have briefly highlighted the key principles of modern methods of measuring internal and international migration.

However, measuring ongoing migration processes in applied contexts in a manner that is *conceptually and methodologically sound*, *mathematically, statistically and demographically consistent*, and *maximally informative*, inevitably involves a range of more detailed further issues, both of a methodological nature and of an operational nature, that are beyond the scope of our foregoing presentation.

For a more comprehensive resource on the current state of the art in the measurement of both internal and international migration processes, the reader is referred to Xu-Doeve (2006).

This work deals in depth with many of the key topics, ranging from empirical context, current practice, demographic theory, concepts and operational definitions, data specifications, data sources and data collection procedures, demographic measurement, and the estimation of and adjustment for underenumeration, to numerical analysis, data processing, and appropriate information systems design.

References

De Witt, Johan (1671) *Waerdye van Lyf-renten naer Proportie van Losrenten (The Value of Life Annuities Relative to Redemption Bonds)*. Agenda item in: Resolutiën van de Heeren Staten van Holland en West-Friesland (Resolutions of the Assembly of Deputies of Holland and West-Friesland)

Du Pasquier, L G (1912) Mathematische Theorie der Invaliditätsversicherung. In: *Mitteilungen der Vereinigung Schweitzerischer Versicherungsmathematiker* 7:1-7

Du Pasquier, L G (1913) Mathematische Theorie der Invaliditätsversicherung. In: *Mitteilungen der Vereinigung Schweitzerischer Versicherungsmathematiker* 8:1-153

Keyfitz, N (1970) Finding Probabilities from Observed Rates or How to Make a Life Table. *The American Statistician* 24:28-33

Keyfitz, N (1977) *Introduction to the Mathematics of Population. With Revisions*. Addison-Wesley Publishing Company

Keyfitz, N and H Caswell (2005) *Applied Mathematical Demography. 3rd Edition*. Springer

Pressat, R (1983) *L'Analyse Démographique : Concepts, Méthodes, Résultats. 4th Revised Edition*. Presses Universitaires de France

Rees, P (1984) Does It Really Matter Which Migration Data You Use in a Population Model? Working Paper 383, School of Geography, University of Leeds

Rees, P and A G Wilson (1977) *Spatial Population Analysis*. Edward Arnold

Rogers, A (1975) Introduction to Multiregional Mathematical Demography. John Wiley & Sons

Shryock, H S, J S Siegel and associates (1971) *The Methods and Materials of Demography. 2 Volumes*. US Bureau of the Census

Siegel, J S and D A Swanson, eds (2004) *The Methods and Materials of Demography. 2nd Edition*. Elsevier/Academic Press

United Nations (1970) *Manual VI. Methods of Measuring Internal Migration*. United Nations

United Nations (1997) *Principles and Recommendations for Population and Housing Censuses, Revision 1*. United Nations

United Nations (1998) *Recommendations on Statistics of International Migration, Revision 1*. United Nations

Willekens, F (1994) Monitoring International Migration Flows in Europe. Towards a Statistical Data Base Combining Data From Different Sources. In: *European Journal of Population* 10:1-42

Xu-Doeve, W L J (2006) *Methods of Measuring Internal and International Migration*. ANRC Publishing (ISBN-13: 978-90-8802-001-8)

Xu-Doeve, W L J (2007) The Applied Measurement of Migration. Paper invited by the United Nations Statistics Division for presentation at the 56th Session of the International Statistical Institute (ISI), Lisbon, 22-29 August 2007

CHAPTER 2

Chapter 2

Modern Methods of Measuring Internal and International Migration:

A Synoptic Overview of Concepts, Data, Data Sources and Data Processing

This chapter was first prepared in 2008 as a background document on methodology for the United Nations Expert Group on the Use of Censuses and Surveys to Measure International Migration.

ABSTRACT

In recent years, new and powerful methods of *measuring internal and international migration* have emerged. As it turns out, they also represent the conceptual, operational and analytical convergence of the hitherto often distinct approaches to measuring internal and international migration. Based on rigorous methodological principles, these modern methods unambiguously define:

(1) the *optimal specifications of migration data* to be collected
(2) which *data sources* best to use
(3) the *most informative methods of measurement* of *intensities*, *flows* and *stocks*, and of their *dynamics over time*

These new methods of measurement result in *empirically comprehensive and fully-detailed information* on ongoing or historical migration processes.

This chapter centres on migration data. The focus is in particular on the most important *implications* of modern methods of measuring migration processes for *data specifications*, for *data sources* and *data collection*, and for *data processing*. The chapter presents a synoptic overview of these topics, *summing up the main points* and *outlining the basic principles* in a non-technical manner. It centres on key results for *direct practical implementation and application*.

KEYWORDS

International Migration, Internal Migration, Migration Data, Migration Data Collection, Migration Data Sources, Population Register, Population Census, Population Survey, Data Processing, Methods of Measuring Migration

Copyright © 2008 William L J Xu-Doeve

TABLE OF CONTENTS

1 Introduction

In recent years, new and powerful methods of *measuring* *internal and international migration* have emerged.

At the same time, they also represent the conceptual, operational and analytical *convergence* of the hitherto often distinct approaches to measuring internal and international migration.

Based on rigorous methodological principles, these modern methods unambiguously define:

(1) the *optimal specifications of migration data* to be collected
(2) which *data sources* best to use
(3) the *most informative methods of measurement* of *migration rates*, of *migrant flows* and *stocks*, and of the *dynamics over time* in these rates, flows and stocks

These new methods of measurement result in *empirically comprehensive and fully-detailed information* on ongoing or historical migration processes. Measurement results can include, for example, detailed information on:

(1) migration *rates* (migration *intensities*; *propensities* to migrate), and their *dynamics over time and age*
(2) migrant *flows* and migrant *stocks* in absolute numbers, and their *dynamics over time and age*
(3) migration *trajectories* (geographical routes or paths taken by migrants), including for example also circular and return migration

(4) *temporary* and *short-term* migration versus *long-term* migration
(5) *frequent* migration
(6) *estimates of* and *adjustments for* the often highly *incomplete enumeration* or *registration* of recent migrants
(7) clear, well-defined and consistent approaches to more *in-depth analyses* and to *explanatory studies* of migration behaviour

These new methods of measuring migration have rigorous foundations in mathematics, they are based on firm methodological principles, and they are demographically fully consistent. Yet, as we have seen in chapter 1, their practical operational application is remarkably simple, straightforward and transparent. Given correctly specified data, the measurement of migration processes is a procedure that can easily be made into a standard routine.

The present chapter focuses on *migration data specifications* (data definitions), *migration data collection* and *migration data processing*. It complements chapter 1, where the focus is on *methods of measurement*.

Chapter 1 was presented earlier as Xu-Doeve (2007), and a condensed version, Xu-Doeve (2008), is also available in the Proceedings of the 56th Session of the International Statistical Institute. A more comprehensive and in-depth treatment of the material of chapters 1 and 2 can be found in Xu-Doeve (2006). This 2006 book also contains full details and proofs of the underlying mathematical theory.

The present chapter 2 on migration data *sums up the main points* and *outlines the basic principles* in a non-technical manner. It has been prepared with readers who have a limited background in strict formal demographic analysis foremost in mind. Its principal aim is *to assist statistical agencies, universities and other research institutions involved in migration data collection and analysis*.

The chapter centres primarily on key results for direct practical implementation and application, rather than on the fundamental considerations that lead to these results. In other words, it emphasizes *practical methodological and operational principles*, and their *implications and consequences for data specifications, for data collection, and for data processing*.

For the benefit of the reader, the penultimate section, section 7, briefly sketches the underlying *basic principles of applied measurement*. However, for a full understanding of the underlying rationale and the mathematical and methodological logic of the argument, the reader is referred to chapter 1 or, for more detail, to Xu-Doeve (2006).

The remainder of this chapter is organized as follows:

We begin in the next section 2 with a brief summing up of *key concepts and definitions*. Then, in section 3 we outline some of the most important *implications for the collection of data* on internal and international migration. In section 4, we move on to analysis of a more explanatory nature. In particular, we discuss how the use of co-varying personal attributes (characteristics) ties in with fundamental *conceptual and methodological problems that are commonly encountered* in applied migration data collection and measurement.

Section 5 discusses the three most important *data sources* that may be considered for the establishment of a benchmark comprehensive image of migration processes: *periodic population censuses*, *continuous population registration systems*, and *probabilistic sample surveys*. It is shown that, in most countries, the periodic population census will be the data source of choice. Next, this section 5 briefly deals with some of the major *limitations of the population census* as a source of data on migration processes, and it points towards key *approaches to deal with these issues*. This is followed by a brief section 6, highlighting some of the most important *implications* of the measurement of migration processes for *census data processing*.

In section 7 we present a bird's eye view of the actual *methods of measurement and analysis* themselves. This serves, first, to put the material presented in this chapter 2 in a more rigorous methodological perspective, and, second, to build bridges to the actual applied operational measurement of migration processes presented in chapter 1.

Finally, in section 8, we conclude with a discussion of *selected key related publications* that deal with methodological aspects of the collection, processing and analysis of data on internal and international migration.

2 Concepts and Definitions in the Analysis of Migration

A ***migrant*** is a person who has experienced one or more *migratory events* during his or her life time. A ***migratory event*** or ***migratory move*** is a ***change in a person's place of usual residence***.

Thus, the elementary *key to measuring, characterizing, interpreting and understanding migration processes* -- migration intensities, migrant flows and stocks, migration trajectories, and so on -- is the ***recording of migratory events (migratory moves) as these occur in the life histories of persons as time and age progress***.

In other words, in its essence *migration data collecting* is simply a question of *tracing the lives of persons* and *recording the migratory events* (migratory moves) of interest, *specific by time* (or, interchangeably, by *age*) of occurrence and specific by the place of residence immediately prior to the move and by the place of residence immediately subsequent to the move, that is, *specific by* the *places of origin and destination* associated with each particular move.

A migratory event is an ***internal move*** if the change in the place of residence does not involve the crossing of an international border. Otherwise it is an ***international move***.

A person having experienced an internal move is called an ***internal migrant*** in respect of that move. A person is called an ***international migrant*** if the move in question is an international one.

It is important to underline that a person is *operationally defined* as a migrant on the basis of *observed migratory events*. Common practice in migration studies frequently runs counter to this principle: Often, persons are operationally defined as migrants on the basis of other ***attributes*** (*characteristics*, *properties*) *taken as* **proxies** *for migratory events, events which themselves have remained unobserved*.

Well-known examples include proxy variables such as an observed difference between the current and an earlier place of residence, or a nationality that differs from the one of the country of current residence.

The quality of such proxies as indicator variables of migratory events varies from case to case, but the most important issue is precisely that these proxy variables all have in common that they leave the actual migration process itself unobserved.

Consequently, even the most elementary aspects of migration processes, such as migration intensities, true migrant flows and stocks, and migration trajectories, cannot be ascertained with certainty. Neither is it possible to establish any margin of error around the findings which quantifies the degree of this uncertainty.

For these reasons, any interpretation of such *proxy migration data* in terms of actual ongoing or historical migration processes necessarily has to remain problematic, both methodologically and empirically. It is a key issue in the applied measurement of migration to which we shall return in section 4, below.

3 MIGRATION DATA AND MIGRATION DATA COLLECTION

For the collection of a data set that enables the analyst to capture a comprehensive image of ongoing or historical migration processes, there are generally only two alternative data sources that meet the necessary requirements, namely a *continuous system of population registration* in which changes of residence are recorded, and a *periodic population census*.

Often, *sample surveys* are mentioned as a third data source. However, migration processes are specific *at least* by age (cohort), by timing of the moves in the life histories of cohort members, and by geographic trajectory involved in the moves. Consequently, probabilistic *sample surveys* drawn from the general population, even when sophistic designs are used, cannot easily reproduce such processes within reasonable boundaries of sampling error, that is, within acceptable margins of uncertainty resulting from the randomness underlying unit selection.

Further, few countries have continuous registration systems adequate for the purpose. This, therefore, usually leaves the population census as the only suitable source. Accordingly, in this section we shall focus principally on population censuses as our benchmark source of migration data. We shall return in more detail to the issue of data sources for the measurement and analysis of migration processes in section 5, below.

In a population census, the *data on the migratory moves* experienced by respondents during their respective life histories -- the occurrence of any moves, their geographical trajectories in terms of origins and destinations, and their timing --, are obtained through *retrospective questioning* on past behaviour.

The precise formulation of the questions used may differ somewhat from national statistical office to national statistical office and from census to census. However, in their essence these questions on the migratory events experienced are framed so as to capture ***current and previous places of usual residence***, together with the associated ***durations of residence*** in each of these places.

In the next sections 3.1 - 3.4 we shall explore such questioning in depth. The aim is to obtain migration data sets that are *maximally informative*, and adequate for the comprehensive measurement and analysis of ongoing and historical processes of internal and international migration.

In the applied practice of population census taking, methodological and operational concessions are often made, however, which reduce the informative value of the resulting migration data sets. This will be a topic of special attention in these next four sections.

We note that in many censuses *fundamentally different approaches to questioning* on past migration behaviour are chosen. This will be the subject of section 4, where we shall discuss the most common of such alternative lines of questioning and their implications for the measurement of migration processes.

3.1 Data Collection and Data Analysis -- Two Distinct Roles

Using population censuses as the primary source of migration data has a significant operational implication: The full basic picture of the nature and dynamics of actual migration processes is obtained by resorting to a data collection mechanism that by its very nature has *multiple purposes* and that is *not specifically* designed for the collection of migration data.

From the point of view of the measurement and analysis of migration, one important consequence is that in practice there inevitably arises a significant *separation and division* between the *role of data collector* on the one hand, and the *role of data analyst* on the other.

The task of the data analyst is to provide evidence-based answers to specific questions on migration posed for specific purposes by policy makers, by executives, by managers, and by all those involved in one role or another in monitoring, evaluation and assessment.

For the data collector, that is, normally, the national statistical office (NSO), it is likely to be difficult to anticipate the objectives, scope and depth of all such questions, and in advance to tailor the data collection effort accordingly.

Consequently, it is of fundamental importance that the the NSO *in its role as the data collecting agent* *does not unnecessarily restrict, pre-empt or prejudice* *the role of the data analyst* in terms of

(1) *a priori selection* in the *observation of migratory events*;
(2) *pre-determined operational definitions of concepts*; and
(3) *pre-defined data groupings and aggregations*.

This principle is especially significant in the present context of the measurement of migration, given the many dimensions of an empirical process inherently as complex and variable as migration, and given the many angles from which such processes can usefully be studied.

In other words, it is important that the NSO collects such data that the migration *analyst* is allowed maximum scope and flexibility so as to be able to answer specific questions that may be raised on ongoing or historical migration processes.

Any inappropriate or suboptimal decisions taken at the stage of data collection can all too easily impair any subsequent meaningful and useful analysis of migration processes, or even render such analysis impossible.

This issue will therefore form a central thread in the remainder of this section. It is a perspective that usefully allows us to compare and contrast common practice in population censuses with the *requirements* and the powerful *opportunities and scope* of modern event-based approaches to measuring migration processes.

Data on migratory moves may be collected ***selectively***. In other words, *in the data collection process* the *a priori choice* may be made *not to observe all moves* experienced by respondents during their respective individual life histories.

As regards both *international* and *internal migration*, for example, it is common practice in population censuses that, if such data are recorded at all, only data for *the most recent move* (last move data) are recorded.

However, many persons experience more than a single move during their life histories. Well-known examples are, for instance, temporary and short-term migration, frequent migration, seasonal migration, circular and return migration.

Clearly, therefore, observing *at least the two most recent moves* would already be considerably more informative and preferable. However, the obvious aim should be the establishment of yet more complete migration life history records:

In this context it is worth noting that the mathematically and methodologically consistent measurement of other demographic processes, such as fertility and mortality, is formally similar to the measurement of migration processes. While the event of death occurs only once in the life history of a person, the event of giving birth can occur more than once -- similar to the event of experiencing a migratory move. In the study of fertility it would be considered inconceivable and wholly unsatisfactory to limit empirical observation as a matter of course to not more than one single birth during the life history of each person. Yet, in the collection migration data through population censuses, this has become the de facto established norm and accepted practice.

Clearly, limiting the observation to at the most a single event in the life histories of persons is as unsatisfactory in the case of migration as it would be in the case of fertility. We shall briefly return to this issue from a different perspective in section 3.3, below. And later, in section 5, we shall discuss practical approaches to actually realizing the observation of more than merely the last move (the most recent move) in population censuses.

As another example of a priori selectivity in the observation of migratory events, for the study merely of *international migration*, all internal moves might be disregarded in the data collection process. While disregarding all internal moves might seem an odd choice for an NSO to make in a population census, it is, in fact, not uncommon in actual census taking.

> The following situation is a case in point: For international migration, direct questions are asked on country of previous residence and on the duration of residence in the present country since arrival. Here, therefore, the migratory moves themselves are recorded as they have occurred in the life histories of census respondents. For the measurement of internal migration, on the other hand, a question is instead asked on a proxy variable which is taken as an indicator of migratory behaviour, without, however, recording the actual migratory moves themselves. A common example is the use of a question on the place of residence 1 or 5 years prior to the enumeration. Such a question merely establishes that the place of residence 1 or 5 years ago was different from the current one; but it does not actually record any migratory moves themselves, nor their timing. (We shall return in more detail to this type of questioning in section 4, below.)

It will be clear that the more selective the recording of migratory moves in the data collection process is, then correspondingly the greater the a priori restrictions that are placed on the subsequent ability comprehensively to measure and analyse actual true ongoing or historical migration processes.

It is also common practice in population censuses that the *scope*, *depth* and *precision* of the information gathered is ***a priori narrowed down*** *in the data collection process*, namely by employing both *concepts and operational definitions* and *operational data collection practices* that are *unnecessarily restrictive*.

The extent to which this places limitations on the subsequent measurement and analysis of migration is often less immediately obvious, because this ties in closely with some of the deeper elementary methodological aspects of the measurement of migration processes (see also section 7, below; for more details see chapter 1). We shall highlight the most important implications and considerations for population censuses. They are grouped in the following three subsections 3.2 - 3.4, dealing respectively with the operationalization of the concept of *usual residence*, with the operationalization of the concept *place of residence*, and with the operationalization of the concept *timing of events* (timing of migratory moves). If other data sources than population censuses are used, then the essence of the matters discussed in these subsections 3.2 - 3.4 similarly holds true.

3.2 Usual Residence

Recall from section 2 that a migratory move is conceptually defined in terms of a change in the place of usual residence of a person. In operationalizing this notion, the question immediately arises: How do we interpret the concept usual residence, and when do we speak of a change in this usual residence? Here, the following two operational principles apply:

First, while collecting data, a place of residence should *always* *be considered as usual* *except when* the following two conditions are *both* satisfied: The *stay is short*, that is, *less than one month*, *and* the *purpose of stay* is clearly not associated with an actual change in the place of usual residence.

Typical examples of such purposes of stay that are not associated with an actual change in the place of usual residence, include: *recreation, tourism and holidays*, *business travel*, *temporary visits to friends and relatives*, *temporary medical treatment*, and *religious pilgrimage*.

Possibly this list could be extended with similar further purposes. With a view to ensuring the international comparability of migration data, it is recommended that an exhaustive international standard list of such purposes be compiled and used.

Second, in the data collection process, a *place of usual residence* must be interpreted and recorded *as it actually is*, that is, *on a de facto basis*, and *not on a de iure basis*.

These two principles are essential if one wishes to allow the study of important issues such as, for instance, *temporary*, *short term* and *seasonal* migration.

Further, these two operational rules *allow and enable the analyst to decide*, depending on the context and purpose of the study undertaken, how appropriately to distinguish between long-term ("permanent") and short-term ("semi-permanent" and "temporary") residence. These principles leave the analyst the opportunity to decide which observed moves, if any, to exclude from a particular study as not relevant given the scope and objectives of the analysis in question.

It is common to see that a *minimum residence duration* -- such as at least 6 months, or at least 12 months -- is built into the data collection process as an operational criterion to determine whether or not a place of residence is a *usual* one. If the actual residence duration has remained below this limit, then the residence is not regarded as usual. Consequently the change(s) in place of residence involved, that is, the move(s), are interpreted as not migratory, and the event(s) are not recorded.

This is a clear case of the undesirable situation where the data collector pre-emptively takes on the role of the analyst, and prejudicially and a priori eliminates the observation of what in many studies of migration may well be regarded as significant migratory events.

Thus, clearly, the criterion is not only arbitrary, it is also unnecessary and uncalled-for. When residence durations associated with all actual moves are properly recorded (see also section 3.4), then the analyst is left with every freedom to draw any distinctions between persons on the basis of the criterion of residence duration if this is deemed appropriate: any classification scheme remains an option. The *full richness of actual empirical migratory behaviour then is preserved* in the observation process, and it allows the analyst to let the data speak the story in its entirety.

Importantly, also, if for the purpose of data collection a place of residence is a priori operationally defined as usual on the basis of some such arbitrary minimum residence duration criterion, then it should come as no surprise that later the analyst cannot but find that *all migration has stopped 6 or 12 months prior to the date of data collection*.

An issue closely related to the concept of usual residence is the question whether in the collection of migration data a *de iure or a de facto* approach should be adopted in determining what should in fact be recorded as usual residence.

If a *de iure* residence criterion is built in into the data collection process, then many actual migratory events may remain unobserved for the simple reason that a person has merely *not changed residence in administrative or official terms*.

The resulting discrepancies between actual true migration processes on the one hand and recorded migration based on a de-iure definition of places of usual residence on the other, can further increase significantly if respondents are made to understand that only *formal* or *official* places of residence must be given:

If there is even just the *perception* among respondents that authorities might disapprove of their actual place of residence now or in the past, then *deliberately erroneous answers* are also likely. As personally intrusive instruments to collect data on empirical reality, population censuses perform best when they are perceived by respondents as confidential, safe and independent of any government, legal, tax, and other authorities. The enforcing any explicit official de iure interpretations in the questioning of respondents contributes to negatively affecting such perceptions.

We conclude this section with a note on the use of the concept of *intentions* and of the concept of *motives* in the collection of migration data. Such concepts are sometimes advocated as suitable to determine whether or not residence is to be interpreted as usual or not.

The measurement of migratory moves is based on the factual occurrence of moves. In the measurement per se of migration processes, that is, in the measurement of migration intensities, flows and stocks, issues such as any *intentions to stay*, any *intended durations of stay*, or, for that matter, any *other intentions*, do not play a role.

Similar to the measurement of other demographic processes, such as the measurement of mortality rates or fertility rates and the associated absolute numbers of deaths or births, the measurement of migration processes is behavioural: It fundamentally centres on the observation of actual migratory events as these occur in the life history of cohort members. Consequently, the essence of migration data collection, therefore, is the recording of de facto migratory moves.

This behavioural approach is elementary and essential for the measurement of migration processes. However, methodologically it is also often preferable in the subsequent analysis and interpretation of the findings.

> For example, as we just saw, the analyst will usually prefer to determine whether a move is to be qualified as temporary or permanent directly on the basis of the actual length (duration) of the stay, not on the basis of any stated intentions. In making such a distinction, the criterion for this duration, whether it be 1, 3, 6, 12 months, or some other period, is essentially arbitrary. It is the analyst who, given the purpose and scope of the study, has to make a reasoned choice, here. The knowledge of intentions is unnecessary for such a classification. In addition, in a population census, or, for that matter, in most other common data collection efforts, it will in practice normally remain unobserved whether or not the stated intentions are actually realized. Factually occurring migration behaviour, on the other hand, is directly observable. This also makes it more objectively comparable across cohorts, populations and over time.

Similarly, the *purpose of stay* or any *other motives or reasons* play no role in the measurement itself of migration processes. There is only one exception.

This exception is made in the case of *residence durations under one month*, as described above at the start of this section 3.2. However, in the case of these very short durations of stay, it merely serves to distinguish between travel or similar movement without a change in usual residence on the one hand, and migration on the other: It is necessary so as to allow the proper identification and measurement of temporary and very short-term migratory moves and of very recent migratory moves.

While the measurement of migration processes itself is not a function of stated intentions or of stated motives, this is, of course, not to say that therefore such intentions and motives are irrelevant in the study of migration. Depending on the objectives and scope of the analysis undertaken, a statistical explanation and a deeper understanding of observed migration intensities, flows and stocks may well be served by a knowledge of stated underlying intentions or motives. However, such variables are co-varying characteristics in the analysis, rather than instrumental variables in the measurement itself of the migration process.

We conclude this section by noting that if in what follows we speak of a place of residence, then this should be interpreted as a place of residence in the sense of migration, that is to say, as a place of usual residence as defined above in this section 3.2.

3.3 Places of Residence

Among the many key aspects of migration processes that are of interest for measurement and analysis, one of the most basic and elementary, of course, is the *relocation of population members*, and the resulting redistribution of populations. In order to facilitate adequate insights in this, it is fundamental that such relocation is captured in the data collection process in a manner that is as rich in information as possible.

Formulated somewhat differently, a matter of primary concern in charting and understanding migration processes is the ability adequately to measure, study and interpret precise *geographical migration trajectories* (migration paths or routes) and *associated migration flows*.

This ability is secured only if *in collecting migration data* current and previous *places of residence* are *observed and recorded* *with sufficient geographic detail*.

At first sight this may perhaps seem an obvious and rather trivial statement. However, the practice of population census taking often differs substantially: The incorporation of a priori regional (geographical) classifications *at the stage of data collection* is common. In data recording and/or in data coding, individual places of residence are often already aggregated (grouped) by some set of pre-defined regional criteria.

In the case of *international migration*, places of origin of respondents who have immigrated, for instance, are commonly grouped in countries.

For most analysis, this may be quite adequate. However, for major countries of origin, a finer disaggregation of places of origin may well be worth considering with a view to enabling more informative measurement and analysis. This could involve, for example, differentiating at least between major urban agglomerations, rural zones, and so on.

In the case of *internal migration*, places of residence are also usually grouped regionally. Commonly, a set of pre-defined operational regional subdivisions is used, such as, for instance, localities, districts, major or smaller civil divisions, urban and rural areas, enumeration areas, and so on. In the resulting census micro data, the effect of such a priori classification is usually most pronounced on places of previous residence, but micro data on places of current residence may also be affected.

However, such pre-defined regions may well in fact disguise significant internal population flows. Further, these regions may also conceal important internal heterogeneity. For example, it quite possible that some parts in such a given region may be an important, and possibly highly selective, origin of or destination for migrants, whereas the remainder is much less affected. This is, for example, typically the case within many urban areas.

The empirical significance of this issue of the prior classification of places of residence in broader regions is reinforced by the well-known fact that often the highest migration intensities are associated with moves over relatively short distances. Consequently, the use of crude regional aggregates to group places of residence as early as at the stage of data collection may effectively eliminate the observation of precisely the most important migration flows.

There is also a strongly self-reinforcing circular argument between opting to use broad a priori regional groupings in recording places of residence on the one hand, and the more or less universal practice in population censuses of opting to limit the observation of migration behaviour at best to the most recent migratory move only: The a priori regionalization used can easily obscure so much of actual migration behaviour that the number of persons for whom an earlier move can still be observed, becomes simply too small to be worth the effort. This then is an *operational artefact*, and it does not necessarily reflect actual empirical reality.

In reality, of course, the occurrence of multiple migratory events to persons can be common and significant. Well-known instances include, for example, temporary and short-term migration, frequent migration, seasonal migration, circular and return migration, and so on.

Clearly, if NSOs regard the collection of data on migration processes as a matter of importance, then it is elementary that population censuses result in migration data sets which encompass the breadth and depth of the actual migration experience of the population.

This requires *questioning beyond merely the most recent move*, together with an *adequate geographical resolution* (detail) allowing actual migration behaviour to be borne out.

In summary, all prior geographical groupings of places of residence in the data collection phase *subsequently deny the analyst the possibility to study* those elements of ongoing and historical migration processes that have been *deliberately and irrevocably kept hidden from view* as a direct consequence of this very practice of place of residence grouping in the empirical observation, recording and/or coding of the census micro data.

One argument that is sometimes put forward, is that any intra-regional movement that has thus been made unobservable, is -- by implicit definition -- not actually migration but *residential mobility*.

However, residential mobility is a concept that itself has no special basis in the mathematical theory underlying migration processes. It is useful to explore this matter, because it touches on an important methodological issue in data collection and in the measurement and analysis of migration.

Mathematically, a move is merely an *abstract event* that may occur either never or once or more times in the life history of a person. In order to associate such an abstract event with an empirical interpretation, a suitable *operational definition* is required. Methodologically, we proceed as follows:

Given the scope and objectives of the study to be undertaken, the analyst begins by defining the geographical areas among which migration is to be measured, analysed and interpreted. Then a *migratory move* is *operationally defined* as the event of a change in the place of residence *if and only if* this change in the place of residence involves a relocation from one such geographical area to another. It is precisely for this reason that such geographical areas are therefore called ***migration-defining areas*** (MDAs).

From the point of view of the measurement of migration, the analyst is entirely *free to define these areas*, both in terms of their number -- two or more --, and in terms of their geographical demarcation. The only condition is that MDAs be non-overlapping. But they need, for example, not be exhaustive nor geographically contiguous. The choice in this matter should be determined exclusively by the scope and objectives of the study to be undertaken.

Thus, it is important to understand that, for the purpose of any particular given study, only inter-MDA moves are considered. By the above operational definition, intra-MDA moves fall outside the scope of study. If the latter is unsatisfactory given the scope and objectives of the study in question, then, in other words, by definition the analyst has simply misspecified the MDAs.

To put this in perspective, it can in fact be entirely appropriate that the migration-defining areas be defined at the finest operational level possible, namely as the set, or a subset, of all available individual places of abode, that is, of all available places of residence in the overall total area under study.

The use of a geographically relatively fine specification of MDAs is typical in what is sometimes called the study of residential mobility. Residential mobility, furthermore, is commonly understood to apply at some geographically restricted local level only. However, methodologically, as far as the measurement is concerned, the distinction between migration and residential mobility is arbitrary and artificial. At the most, it is simply a practical matter of the operational definition of the set of MDAs considered.

However, the important point is this: If such a distinction between migration and residential mobility might be appropriate in view of the scope and objectives of the migration study to be undertaken, then it is the role of the analyst operationally to define precisely which moves are to be excluded from the analysis as non-migratory. The national statistical office (NSO) in its role as the data collector should not play a prejudicial role, here.

The *specification of migration-defining areas* (MDAs) is, in fact, a crucially important general issue in the measurement of migration. In many cases the migration analyst will be confronted with questions that require the definition of migration-defining areas (MDAs) *tailored specifically* to meet the purpose, scope and objectives of the particular study to be undertaken.

Any regional definitions enforced as a consequence of unnecessary or avoidable decisions made earlier in the data collection, coding, input and processing phases, merely reduce the ultimate value of the costly data collection effort as a means to obtain relevant information on ongoing and historical migration processes.

Clearly, the coarser any pre-defined regional aggregates that are used in the data collection, coding, input and processing stages, then the less informative any subsequent measurement and analysis of migration can be. Best, of course, is that any such a priori geographical grouping be avoided altogether:

First, instead of place of residence grouping, precise *geo-referencing* in the *data collection* process is preferable, and this maximally empowers the analyst. To the extent possible, it is recommended that Global Positioning System (GPS) co-ordinate data are used. Suitable alternatives, where feasible, are, for example, complete address data and/or postcodes.

Second, it is essential that such geographic precision as is available in the raw collected data is *not subsequently lost* in *data coding, input or processing*, and that the ability is maintained to *make the original geographic detail available* later *on request* for migration measurement and analysis *in the form of migration-defining areas* (MDAs) that are *tailor made* to suit the scope and objectives of the particular empirical study to be undertaken.

3.4 Durations of Residence or Timing of Moves

We recall once more that the measurement of migration processes fundamentally centres on the observation of migratory events (migratory moves) in the life histories of persons. In the previous sections, we have discussed the operational implications for data collection of the interpretation and observation of migratory events. In the present section, we shall explore the issue of the *position of such events in the life histories of persons as time and age progress*, and the consequences for migration data collection.

In other words, here we shall be concerned with the *timing of the migratory events* experienced by respondents during their respective life histories, that is, the *date of each move*, or, equivalently, the *durations of residence until a move and between successive moves*.

The key issue for migration data collection is that these event timings must be *observed and recorded with adequate time resolution*, that is, on a sufficiently precise time scale. This requirement is a direct implication of the underlying mathematical theory of migration and of the associated operational methods of measurement which logically and directly follow from the mathematical theory. Refer to section 7, below, for a brief non-technical overview of basic methods. For a more in-depth treatment, see chapter 1.

Observation and recording of the timing of migratory events precise down to *years and months* is *adequate*, and a precision down to *years and quarters or seasons* is still *acceptable* for the informative measurement of migration processes. A time precision expressed in *years only* is, however, the *absolute minimum* so as still to allow any meaningful measurement procedure.

Any cruder classification of migratory event timings effectively renders any meaningful measurement impossible.

It is, for instance, quite common practice in population censuses to ask of respondents that they indicate their durations of residence in three broad classes, such as, for example, $[0, 1)$, $[1, 5)$, $[5, \infty)$, the first class being 1 year wide and ranging from 0 to 1 years in duration, the second being 4 years wide and ranging from 1 to 5 years, and the third ranging from 5 years in duration and upward. Such an a priori classification at the stage of data collection effectively denies the analyst subsequently all possibilities properly to measure the properties that characterize the nature and dynamics of any ongoing and historical migration processes.

In a population census, retrospective questioning is used to obtain information on migratory event timings. The response to such questioning can exhibit a tendency towards digit preference, resulting in event "heaping". This, however, should not be taken as an argument by national statistical offices to use broad residence duration categories in questioning respondents. There are two reasons for this:

First, in this instance the broad categories merely serve to disguise the effect of a tendency towards digit preference among respondents. However, instead of hiding this tendency, the analyst should be left with the opportunity to investigate the raw data, and to explore if they actually do exhibit any signs of heaping in the recorded event timings. If this indeed proves to be the case, then the analyst is able to explore the nature and extent of the phenomenon.

Second, the presence of event heaping in the raw data does not interfere with or preclude the subsequent accurate measurement of migration processes. This is because the actual measurement procedure itself (see section 7, below) is robust in respect of any digit preference and the associated heaping of event timings.

Clearly, it is important that no unnecessary new or further classification of event timing data be *introduced* after the collection of the raw data in any subsequent *data coding, input and processing*. Here the original precision with which the actual timings of migratory events have been observed and recorded in the collection of the raw data, must, of course, be fully preserved.

As discussed, maintaining an adequate level of precision in the timing of migratory events in the life histories of respondents is a requirement that is dictated by the routine mathematical methods underlying the procedure to measure migration processes. Without a thorough understanding of the principles of demographic measurement, however, it is also immediately obvious that the coarser the time units used, then the less informative any analysis of the migration process in question can be. The following two examples serve to highlight this.

Migration intensities (instantaneous migration rates) and the dynamics of a migration process, for example, both, of course, are functions of time and age. The classification of migratory event timings in the life histories of persons in broad categories (time intervals), therefore, has a direct bearing on the ability to identify such migration process characteristics in the empirical data, and on the detail and the accuracy with which they can be measured.

Further, for instance, if event timings are expressed in, or rounded to, years, then key concepts such as short-term, temporary, frequent and seasonal migration will inevitably remain elusive forever empirically.

To put matters in perspective, it is useful to recall that in many countries the timing of, for example, events of mortality is registered precise down to day, hour and even minute, while generally mortality as a function of time and age is considerably less dynamic than is migration.

In the data collection process, a national statistical office can choose to observe either the *date* of each of the migratory moves made, or the *residence durations* until a move and between successive moves. Either line of questioning effectively results in the same information. These two types of questioning appeal to different perspectives in the recollection of respondents, however.

Using both lines of questioning can be used as an instrument on a trial basis prior to the full census to help selecting the best approach to questioning on the timing of past migratory events in a given population.

In addition, such dual questioning can also be deployed as an internal consistency check instrument on a sample of respondents during the actual census to assist in assessing the accuracy and reliability of the response to such retrospective questioning on the occurrence and the timing of past migration behaviour.

4 Collecting Data on Migrant Attributes

We recall from section 2 that the measurement of internal and international migration processes fundamentally centres on the observation of *migratory events* (*migratory moves*). Such events occur to *persons* as time and age progress in the course of their respective life histories. A person who has experienced one or more migratory events is called a *migrant*.

These *migrants* can have many ***attributes*** (*properties*, *characteristics*) *that may be of interest* in the study of migration. For the purpose of data collection, such attributes must, of course, first be suitably operationalized in the form of observable variables. However, for the main argument in this chapter this methodological difference is not a principal issue. For simplicity, therefore, we shall use the concept of attribute and its empirically observable interpretation, one or more operational variables, interchangeably.

Now, properly, the study of any given empirical migration process begins, of course, by *charting* the process itself; that is, it commences with *measurement*, answering questions of the types *what*, *when* and *where*. These questions serve to *capture*, *describe* and *characterize* the ongoing or historical migration process itself as comprehensively, accurately and reliably as possible. It will be clear that it is such measurement which is the central issue of the present text.

However, having charted the process in question, then, of course, further questions quickly arise, such as, for instance, questions of the types *how* and *why*. These latter two classes of questions relate to *interpretation*, *explanation*, and ultimately, when all questions have been answered satisfactorily, to *understanding*.

The observation of *attributes of migrants* forms one of the several possible key starting points for such *in-depth insights*, and for *explanatory* studies aimed at an *understanding of empirically observed migration behaviour*.

From a methodological point of view, these personal characteristics are *correlates of the behaviour of migrants*. Importantly, however, and contrary to widespread belief and practice, they are *not* operational variables which themselves play a role as *measurement instruments of migration processes*.

This is a fundamental methodological issue in migration data collection and in the measurement of migration, and it is this issue with which we shall deal in the present section.

There is one attribute that is of special importance, since it ties in directly with the measurement of migration: Recall once more that measuring migration processes entails the observation of migratory moves as these occur in the life histories of persons. Therefore, an elementary attribute whose observation in migration data collection is obviously valuable, is *age*, that is, the *time elapsed since birth*.

However, we note that, without any loss of mathematical and methodological generality, migration can still be measured fully consistently in populations which are *not disaggregated by age*. This is an important issue, methodologically as well as empirically, and for an understanding of the principles of the measurement of migration, it is useful further to explore this.

For the measurement of migration processes, a population is *disaggregated by migration-defining areas* (MDAs) at a point in time that is taken as the starting point of the analysis. When using population census data, then it is often convenient, but not necessary, to take the date of the enumeration as this benchmark time point. It is usually denoted simply by $t = 0$. The essence of the measurement of migration then centres on the observation of the lengths or intervals of *time* until and between migratory events, that is, the *durations* of residence until and between migratory moves, starting at $t = 0$. Section 7, below, briefly outlines the basic principles.

Clearly, *age runs in parallel with time*, so there exists a one-to-one relation between the progress of time until until a person experiences a migratory move on the one hand, and the progress of the age of the person on the other. Strictly, for the measurement of migration, however, only residence durations need to be known. If the age of persons at $t = 0$ is not known, then the measurement method as outlined in section 7 remains fully valid and intact.

Yet, the observation of the age of persons in migration data collection is highly desirable: It is well known that empirical migration behaviour is very closely associated with personal life cycles and that migration intensities (instantaneous migration rates) are strongly related to age. Knowledge of the age of population members at the benchmark point in time $t = 0$ allows the disaggregation of the population in question not only by migration-defining area, but also by age. That is, this knowledge enables us to differentiate between ***age-homogeneous cohorts***, and to *trace each individual cohort in terms of migratory events experienced as time and age progress*.

If we do not disaggregate a population by age, then we implicitly accept considerable *unobserved age-related heterogeneity* in the migration data. The data set then is a mixed bag containing the undifferentiated migration behaviour of cohorts of all ages, that is, of all cohorts, each in its own phase in its life cycle. Clearly, the information obtained for such internally age-heterogeneous populations will be very significantly less valuable than when age-specific cohorts are used. To pursue the metaphor, the data set cannot but speak mixed messages that cannot be disentangled.

We note that it remains common in some scientific domains, such as, for instance, in economics, to study migration processes using data that are not disaggregated by age. However, the empirical evidence of high degrees of age specificity of migration behaviour is compelling. And generally, therefore, we shall formulate measurement methodology and techniques in terms of age-specific cohorts.

We also note that mathematically in the measurement of migration there lies a validity benefit in using empirical data that pertain to persons who are homogeneous with respect to migration behaviour as time progresses. This is an issue that is explained in section 3.2 of chapter 1.

Given the close relationship between both age and prevailing socio-economic conditions on the one hand, and migration experience on the other, a primary, albeit not a complete, proxy indicator variable of this homogeneity is similarity in age. Such similarity is, of course, obtained by observing cohorts, that is, groups of persons who are of similar age, and whose life cycles, as a consequence, also evolve under similar prevailing socio-economic conditions. This issue, therefore, reinforces the importance in the measurement of migration processes of disaggregating populations in age-homogeneous cohorts.

A second attribute that immediately comes to mind as a potential behaviour differentiator is *sex*. Here, a similar argument holds true. Disaggregation by sex is not a mathematical or methodological necessity for the measurement of migration. However, not disaggregating by sex, too, implies the acceptance of potentially significant unobserved heterogeneity in the empirical data.

In addition to age and sex, there are *many other attributes* that are of potential interest in the in-depth and explanatory study of migration processes. Obvious typical examples include, for instance, educational background, family status, economic activity, employment history and status, income, financial transmittance behaviour, and many other ones.

Motives and intentions, although difficult to operationalize and to observe reliably, may also be of special interest for studies aimed at a deeper understanding of migration behaviour.

Specifically for the study of *international migration*, one may consider yet further attributes, such as nationality / citizenship, formal residence permission, mother tongue and language use, country of birth, parental country of birth, and so on.

The scope of the collection of any additional data on such further attributes is determined primarily by the overall purpose of the data collection effort. Here we merely re-emphasize that on methodological grounds none are essential for the measurement of migration processes; only the recording of age is highly desirable.

Yet, the role of such attributes in the measurement of migration is frequently interpreted differently, however. Not unusually, this is related to the use *either* of a *conceptually* selective or biased interpretation of migration and migrants, *or* of *methodologically* defective data definitions and data sets. Often, these two issues are closely intertwined. We shall briefly discuss two of the most important and common examples.

In the case of studies of international migration, *an international migrant* is often *defined* by such additional attributes as nationality or citizenship. Other attributes commonly encountered as criteria to define a migrant in the measurement of international migration are, for example, country of birth, or residence permit status.

Clearly, such measurement instruments are chosen primarily because they are supposed to be indicative of "foreignness". Here we in fact have two issues, one of operationally defining a concept in restrictive terms, and, second, one of suboptimal measurement methodology and associated data collection practice.

In a moment we shall return to the second issue, namely, that of *measuring* migration *by counting persons with some supposedly migrant-defining attribute*, instead of by observing migratory events.

The first issue, and the one that concerns us here, is the specific use of such *migrant-defining attributes* to operationalize the concept of migrant and, by association, *the concept of migration*.

It is important to realize that from the point of view of measuring international migration, such attributes as citizenship / nationality, country of birth, residence permit status, and so on, are *not definitional*. They *merely further characterize* a person. As we have seen in section 2, a person is *defined as an international migrant simply by virtue of having experienced one or more international migratory events*.

Empirically, this distinction can be highly significant. As we have seen in the first part of this section, it can be very instructive and informative to characterize observed migrants by selected attributes. However, this is not the same as using an operational definition of international migrant that has been selected a priori to imply some supposed kind of alien foreignness.

Such a definition excludes all other international migrants. In many countries, in fact, the majority of international migrants do not belong to some class that may be typified as alien foreigners. The common practice of employing a restrictive definition of the concept of international migrant which by its very nature implies alien foreignness, necessarily leads to a selectively skewed and biased interpretation of actual true international migration. It is a distinction that may be clear to the researcher, but it is also one which is all too easily lost on those using the research, such as policy makers and politicians, the media, interest groups, and others.

The second example relates to the use of attributes in such a manner that it leads to methodologically defective data definitions and associated data sets. The attribute in question is the *place of usual residence of a person at some given point in time in the past*, usually 1 or 5 years prior to the census enumeration.

In this case a *person* then is *defined as a migrant* if the value of this attribute differs from the value of the attribute *current place of usual residence*.

We note that this definition of the concept of migrant, too, differs from the definition given earlier in section 2. However, the issue is not just one of the operational definitions of concepts, but also one of the associated consequences for migration data collection:

Given a set of migrant data based on a comparison of places of residence at two distinct points in time, the measurement of migration then is reduced simply to *counting the numbers of persons so defined as migrants*.

Contrary to a widely held belief, however, such data sets are both *inferior* and *defective* on *methodological* grounds as well as on *empirical* grounds. There are, in fact, several reasons for this.

First, the resulting data do not capture any specific *migratory moves*, the fundamental concept underlying the measurement and analysis of migration processes. Put more precisely, such data do not allow one to identify any underlying migratory moves in the life history of persons, neither their true *number*, nor their true *timing*, nor the true *trajectories* involved.

In addition, if for any person the place of residence at the selected point in time in the past is the same as the current place of residence, and if this person experienced any migratory moves in the intervening period, then this *return migrant* is *erroneously defined as a non-migrant*. The migration behaviour involved remains unobserved, and hence also outside any measurement of migration rates (migration intensities) and of migrant flows and stocks.

In all these respects, this attribute closely resembles the attribute place of birth to which we referred above. The only essential difference is that the time period considered is fixed and identical (usually 1 or 5 years) for all persons, and for most persons shorter than the current length of life.

More generally, the *value of the information* contained in the resulting migration data sets is *seriously and fundamentally restricted*. In the introductory section 1 of this chapter, we summed up *seven elementary categories of information on migration processes*. Importantly, because the elementary building blocks of migration processes, that is, migratory events (migratory moves), have not themselves been observed, it is easily shown that *none of these seven categories of information can be obtained in a manner that is both mathematically consistent and methodologically sound*.

On this topic, we finally note that it is sometimes still argued that the collection of migration data on the place of residence at a fixed point in time in the past has a special value for areas such as, for example, population accounting and population forecasting. This argument, however, is without merit: Using data on migratory moves, and following the measurement procedure outlined in section 7, then, if so desired, such information on the place of residence *at any earlier point in time* can be derived in a routine manner.

We refer to chapter 1 for a more detailed discussion of the various issues involved with migration data centred on the attribute place of residence at a fixed point in time in the past.

To conclude this section, we reiterate a key point: *The measurement of migration*, and thus also *the collection of migration data*, *does* ***not*** *revolve around* the

counting of persons who have been *operationally defined as migrants* ***by some observed attribute***.

Yet, as we noted, this is an approach which is common in applied practice. In the case of *internal migration*, studies based, for example, on counts of persons who lived elsewhere one or five years ago, are numerous. And, in the case of *international migration*, the practice to base analyses on counts of persons who were born in another country or who have the citizenship or nationality of another country, is almost universal.

We note, as a related matter of interest, that the well-known *classical indirect estimation of migration* through residual methods (error analysis) also aims at counting persons (migrants). The key differences here are that these *persons themselves remain unobserved*, *at least as migrants*, and that *only their net numbers* are inferred by indirect methods.

Instead, as discussed, the *measurement of migration*, and thus *migration data collection*, *fundamentally revolves around* the *observation and analysis of* ***migratory events*** (***migratory moves***) as they occur in the course of the life histories of persons.

It is a point that is substantiated by logical mathematical and methodological reasoning. More details are available in section 7 and in chapter 1. However, the fact that one does not properly capture a migration process if one does not capture its essence, its fundamental building blocks, namely, the occurrence of migratory moves, is, of course, also immediately obvious.

5 Data Sources for the Measurement of Migration

For the comprehensive measurement of ongoing and historical migration processes, the three principal data sources to be considered are *random sample surveys*, *continuous population registers*, and *periodic population censuses*. We shall deal with these in turn in this section. In addition, in section 8 we shall briefly discuss an example of the use of *administrative data* in the measurement of international migration.

5.1 Probabilistic Sample Surveys

Probabilistic sample surveying is by far the cheapest of the three main migration data collection methods. It involves the smallest data collection operation of the three, and, at least comparatively, the logistics are simple. Consequently, also, sample surveys can more easily be repeated than, for instance, a full population census. This can make it feasible to realize a higher frequency of observation over the years. Further, sample surveying differs from the use of a population register in that it is a dedicated data collection effort. This, together with the comparatively limited scale of the operation, ensures that data collection and processing quality control is easiest to manage. Importantly, too, as compared to a full population census, adequate staff training for the in-depth probing of migration histories can more easily be provided.

These can be powerful arguments in favour of random sampling. Nevertheless, however, relying on sample surveys for the comprehensive measurement of migration processes is problematic.

While developments and trends in, for instance, mortality are constrained by biological factors and governed to a large extent by the gradual progress in environmental sanitation and personal hygiene and in medical science, migration lacks such firm and steady determinants. Migration behaviour is influenced by a wide range of contingencies, including factors such as cultural, economic, social, educational, political and environmental conditions. These conditions provide challenges, opportunities and threats, and they are filtered through individual perceptions, familiarity, networks and experience. Thus, individual migration behaviour is embedded in and influenced by settings which, apart from cultural factors, tend to be highly changeable.

Empirically, as a consequence, migration is *the least stable and predictable* among demographic processes, from cohort to cohort, from place to place, and over time.

In addition, compared to mortality and fertility, migration is multidimensional: It is characterized not only as a function of cohort and of time and age, but also of geographical space. As a concept, spatial dimensionality has one obvious direct implication: By recognizing *variation* in geographical space, it unavoidably adds additional *heterogeneity* to the empirical observations.

Thus, we have a process that is characterized by an intrinsic tendency to be comparatively little stable and poorly predictable, and a process that inherently adds additional heterogeneity to the observations. These are precisely properties which make it difficult adequately to capture a process through probabilistic sample surveys:

Random surveys drawn from the general population cannot easily reproduce such processes within reasonable boundaries of sampling error, that is, within reasonable margins of uncertainty resulting from the randomness that underlies the selection of units for observation. Consequently, in order to capture the variability inherent in migration processes within acceptable bounds of probabilistic error, the production of an adequate data set on migration processes necessarily requires very large samples.

One could, of course, consider more sophisticated sampling designs. For example, one effective strategy to reduce the sample size which is minimally necessary in order to remain within set bounds of sampling error, is stratification. The goal is to reduce heterogeneity prior to probability sampling by subdividing the sampling universe in sub-universes or strata, each of which is internally relatively homogeneous with respect to migration behaviour. The principle is to divide the total variation in the sampling universe into between-strata variation and within-strata variation. The strategy is to select the strata such that as much as possible of this total variation is allocated to between-strata variation. Since as a result each stratum is internally comparatively homogeneous, this allows smaller sample sizes within each stratum for the same level of sampling error. Clearly, however, a sampling design such as this requires prior information on the presence and whereabouts of such heterogeneity with respect to migration behaviour. In many cases reliable information of this kind just does not exist. And there is an important risk in using unreliable prior information for stratification. If the strata are defined in a manner that, given full knowledge, can be shown to be less than optimal, then for the same overall sample size the resulting sampling error can actually increase significantly relative to simple random sampling.

The inevitable conclusion is this: Barring special circumstances, the inherent uncertainty associated with probabilistic sample surveys as a source of migration data effectively places limits on the ability to study migration processes in two key respects. First, *all but any truly major migration flows will remain uncertain*. Other things equal, this uncertainty will, of course, most seriously affect the less mobile cohorts and the thinner routes. Second, the *freedom of the analyst to define detailed migration-defining areas and to study associated migration trajectories will be seriously restricted*. These limitations are quite fundamental, since they go to the heart of the study of migration.

There is also a practical operational difficulty when one wishes to resort to probabilistic sample surveying for the collection of migration data. That is the issue of the availability of an *adequate sampling frame* from which to draw the individual persons to be included in the survey. This requires the existence of, and access to, an up-to-date and complete list of persons in the country with address or other contact details. Such a list must also be made unbiased in the sense that there are no remaining undetected multiple entries for any person. In most countries, no adequate list will be available.

Often, the only moment that there may be such a list is when it is compiled for the specific purpose of the periodic population census. Under those circumstances, it is better not to separate these two data collection efforts, the sample survey and the full census, but instead to integrate them. Below, we shall discuss ways to approach this with the specific purpose to obtain migration data that are both complete and of high quality.

Alternatively, one could consider using the data collected in a population census as a sampling frame for the separate collection of migration data at some later date after the census. However, both due to migration and due to mortality occurring after the census such a sampling frame ages and it becomes more and more obsolete as time progresses. Consequently, in particular the most mobile amongst the population will be missed disproportionally. When the measurement of migration is the purpose of the survey, precisely this is especially undesirable.

We note that it is becoming more and more common that national statistical offices make *random samples of population censuses* available *at micro-data level* for further research by third parties. Generally, such sample data sets are not designed with the specific purpose in mind to serve as migration data sets. Stratification by migration behaviour, for example, is not normally part of the sampling design.

Consequently, in the case of such sampled census micro-data, the same conclusion holds as does in the case of probabilistic sample surveys: At best, they allow the study of truly major migration flows only. And the ability to study detailed patterns of origins and destinations is seriously impaired.

5.2 Continuous Population Registration Systems

For the comprehensive capturing of migratory moves, the *ideal data source* would be a complete continuous population registration system in which *all changes of residence are adequately recorded in real time*, that is, as and when they occur.

Such a continuous population registration system has the advantage that it allows maximum flexibility in the definition of migration-defining areas (MDAs), down to the finest geographical level possible, namely that of individual places of residence. In applied practice, most population censuses, for instance, do not allow for such fine geographical resolution.

In addition, real-time observation of migratory moves means that the resulting data are not at the same time reflecting the competing force of mortality. This is a practical advantage over data obtained in a population census or in a sample survey through retrospective questioning on past migratory events.

Further, a continuous registration system in which changes of residence are recorded captures all migratory events occurring in the life history of each person. Usually in applied retrospective questioning not all moves are actually captured, either by design of the census or survey or by errors of recollection on the part of the respondents. Especially when probing the more distant past, memory lapses will tend to play a more pronounced role.

Real-time recording of migratory events also results in the most accurate registering of the actual timing of each move in the life history of individual persons. Here, too, a continuous population registration system has the advantage over the capturing of events through retrospective questioning.

However, continuous population registration systems are costly to maintain, and relatively few countries operate them successfully. Further, in applied practice, the above benefits are not always fully realized. For example, where such registration systems exist they are often less than complete and up to date, especially in respect of migrants. Recent moves may not have been reported yet, even if there is a regulatory duty to do so.

An important additional issue is that any unauthorized (illegal) migrants will tend to steer well clear of notifying the authorities of their presence and whereabouts. In this respect, the capture of unauthorized migration, population censuses tend to have a clear advantage. This is especially so if NSOs operate fully independently, that is, at arms length of any and all law enforcement agencies, tax collection bodies, and other authorities. The advantage here requires that NSOs can effectively guarantee the non-disclosure of any recognizable and identifiable personal information. We note that, while actual independence and non-disclosure are essential, it is equally essential that NSOs ensure that they are clearly and universally perceived as such. Any perceived breach at an NSO of these conditions as a safe collector and repository of personal information can undermine and destroy this advantage for many years to come.

A further issue with continuous population registration systems can be that they operate on a de iure basis in such a manner that major discrepancies arise relative to actual migration behaviour. An extreme case in point, for example, is the hukou household registration system operated in China.

It is also worth observing that, while data from continuous population registers may be preferred for the completeness and the accurate timing of their life history event observation, such registers are a poor source of information when it comes to in-depth descriptive and explanatory analysis involving covariates of observed migration behaviour. From the point of view of the study of migration processes, continuous population registers tend to hold very little information other than purely demographic and address variables. Elementary socio-economic matters that will feature in any explanatory study of migration behaviour, such as for instance education and employment, normally fall outside the scope of such systems.

Finally, the traditional primary role of continuous population registration systems is administrative and bureaucratic, rather than statistical. NSOs tend not to be considered as users that should have direct access, let alone priority access. Specifically from our perspective furthermore, these systems have, of course, never been designed with the timely and on-demand retrieval of data on migratory events foremost in mind. In practice, conducting even the simplest of data retrieval queries, even in modern digital systems, proves extremely time consuming and costly. The growing awareness that personal data should remain well guarded against improper access, inevitably adds yet further to the practical difficulties associated with such systems in their role as a basic comprehensive data source on migration processes.

5.3 Population and Housing Censuses

The best alternative data source, and for most countries the only realistic choice, is the periodic population and housing census: A full census is usually the only data collection mechanism in which all persons can be probed in respect of any migratory events experienced. Although it is a time-discrete rather than a real-time observation instrument, through *retrospective questioning* it can provide at least a *periodic comprehensive baseline record of earlier migratory moves* experienced to date by all persons in their respective life histories.

In principle and by its nature a full population census is not selective in terms of past migration behaviour. Here, a census has a clear advantage over *administrative sources* which are often selective on the basis of very specific migrant attributes. To illustrate this, we shall discuss the use of residence permit data for the measurement of international migration in section 8 of this chapter by way of example.

In addition, by being a full enumeration, a population census does not suffer from two key problems associated with surveys based on probability sampling designs. First, random sampling inevitably introduces sampling variability and associated uncertainty inherently connected with the probabilistic or chance nature of object selection. As we have seen, this is a particularly problematic issue in the case of migration. Data from a full enumeration embody no such stochastic uncertainty. Second, as discussed, in many countries suitable and adequate sampling frames from which to draw appropriate probability samples for the measurement of migration do not exist.

This second issue does manifest itself in different guise, however, namely in the census mapping and dwelling identification phase of the census operation. The compilation of census maps and dwelling lists is clearly more costly than the use of some existing sampling frame. However, compared to a survey sampling frame, the adequacy of the resulting census maps and dwelling lists is easier to verify and assure. This is because in a census it is dwellings and households, rather than individual persons, that are selected as the primary initial objects in the observation procedure. In addition, in the process of the actual field enumeration, there is the opportunity to detect any errors and omissions in the census maps and dwelling lists, and to correct these. In the case of random sampling, on the other hand, updating the sampling frame on the fly during the survey phase is not realistically possible: It would necessitate that the samples themselves be redrawn to reflect the changes in the sampling frame. Consequently, any interviews already conducted would have to be discarded.

However, the traditional population and housing census can *hardly* be described as *an ideal instrument* for the collection of data on migration processes. In its basic design, the census is a large self-standing multi-purpose data collection effort, conducted at infrequent intervals, usually some 10 years apart.

Many demands are made on the periodic census as a data collection vehicle from widely varying angles and representing very diverse interests. Effectively this means that there is commonly inadequate scope for the collection of migration data beyond the most recent move. Clearly, this does not result in a comprehensive data set on the actual migration processes that shape nations, regions and cities.

Further, the multi-purpose nature of population and housing censuses and the typically large number of questions included inevitably mean that the specialist migration items are frequently dealt with by non-specialists.

Often, for instance, the senior officers at national statistical offices who have the final say in the matter of the number of questions on internal and international migration and of their exact phrasing, are not themselves experts in the area of the measurement of migration. Compromises are common, and one unfortunate consequence is that all too often methodologically suboptimal questions on migration are included in censuses. And, at least as important, the quality of the actual empirical observation cannot but suffer if the census is administered by non-specialist field staff who are insufficiently familiar with the ins and outs of proper questioning on migration behaviour.

In addition, there are several other issues regarding the traditional population and housing census which, at least indirectly, affect the quality and adequacy of the resulting migration data.

For example, the long intercensal intervals combined with the scale of a census operation can cause severely uneven cyclical workloads at national statistical offices. At the same time, the large time intervals between successive censuses inevitably contribute to a loss of continuity in terms of expertise, experience and practical skills from one census to the next.

Especially in many parts of the developing world, the population and housing census effectively remains a non-routine or even a more or less one-off operation. Frequently, too, the processing of the census turns out to be a long drawn-out affair. And especially the handling custom queries on the final census database, queries tailored to address specific issues such as, for example, those typically encountered in the study of migration, often proves very difficult, time consuming and error prone.

Large intercensal intervals and delays in census processing are also the two main reasons why available census results can hardly ever be described as up to date or timely: In the case of a 10-year intercensal interval, the empirical data are on average over 5 years old. And a lack of timeliness often equates to a loss of relevance.

Evidence-based policy making, planning, management, monitoring and evaluation in areas where migration processes are a significant factor, would clearly be better served if such bottlenecks, limitations and constraints could be overcome, or at least be mitigated.

While the issue of modernizing approaches to census taking is increasingly being addressed, especially in developed countries, the vantage point for doing so rarely is the measurement of migration. Yet, as we have seen, for most countries a population census is the only realistic choice when it comes to attempting comprehensively to capture ongoing migration processes.

The key role of the population census as a uniquely significant data source is, in fact, quite special to the measurement of migration processes. Perhaps surprisingly, the census does not have a similarly unique importance for many, if not for most, of the items that are commonly recommended for inclusion as core and optional topics in population and housing censuses. (In section 8 of this chapter, we list the standard references to recent worldwide recommendations and to specifically European recommendations on this matter of core and optional census topics.) Effectively, these recommendations conceive of the population and housing census as a *comprehensive socio-economic and housing survey*.

As of the 21st century, many of these recommendations on topics to be included in censuses have more to do with a legacy of the past and continuity of tradition than with methodological and empirical considerations. This is not to say that the topics are not relevant. The specific issue is that a *full traditional population and housing census* is *often not the most appropriate instrument of data collection*.

One could even reasonably go as far as arguing that, apart from the measurement of ongoing and historical migration processes, the only truly essential ground for taking a full census is the establishment of a record of the population by basic demographic characteristics and by place of usual residence. This would effectively scale the census back from a comprehensive socio-economic and housing survey to a *periodic population registration system*.

For the many items traditionally also included in the population and housing census there are then supplementary and alternative data collecting mechanisms and sources which actually are more appropriate, more informative, more timely, and more cost effective.

For example, the housing component of the census is much more usefully replaced by a combination of remote sensing, that is, aerial surveying and/or satellite imaging, on-the-ground sample surveying, and other sources, such as records kept by relevant service providers.

Similarly, many items these days commonly encountered in the population component of the census are often much more directly available from other sources, including, for example, administrative records and service providers. For information not thus available, these sources can be complemented with the intelligent use of sample surveying. The basic census enumeration and the housing inventory as described above can then serve as two of the possible sampling frames.

The result would be a *light-weight core census programme*, aimed at collecting only those few classes of data for which a full classical enumeration is the only reasonable option.

When conceived essentially as a periodic population registration system, the effectiveness of the census in terms of completeness and quality would be enhanced if such registration is not merely seen as a legal duty, but if it also provides the *opportunity of tangible benefits*, such as access to services. In this respect, there are direct parallels with continuous population registration systems.

Clearly, however, linking the census with the opportunity at the personal level to access services is a matter which requires very careful consideration and the balancing of pros and cons from service to service:

In respect of all personal information collected by NSOs we reiterate the overriding importance of maintaining strong and effective walls between NSOs on the one hand, and government and other agencies and agents on the other. The willingness to participate in the census and the willingness to respond, and to respond truthfully, to all questions depends heavily on the ability of an NSO to guarantee that all personal data are securely kept, and beyond the reach of any third parties. This is not only a matter of regulation and fact, but equally one of perception and trust.

As far as migration is concerned, this is an issue that is especially important if one wishes to capture data on migrants who feel insecure about the possible consequences of disclosing their status. These can be significant groups, and the systematic and comprehensive capture of data on such such specific migrant categories can be extremely difficult otherwise.

Deploying such a much more diversified, creative and versatile approach to the collection of information also allows for a better distribution of the workload over the years. At the same time, the task of the national statistical office becomes methodologically more varied, enhancing career perspectives and encouraging the development and retention of expertise.

In addition, a significantly smaller and more focused set of questions in the census itself will lead to better quality of observation. Furthermore, it facilitates better overall process quality control, and it will lead to a faster overall process turn-around.

Also, a true light-weight census can much more easily be repeated at *5-year intercensal intervals*, rather than at the currently more usual 10-year intervals.

This higher frequency of census taking results in improved continuity of skills, experience and expertise at national statistical offices, as well as in yet further improved operational work flow, operational continuity and operational routine. All such factors, too, add yet further to overall quality.

This combination of a higher census repeat frequency, higher observation and process quality, and faster turnaround will result in data that are more reliable as well as more timely and up to date, and thus in empirical information that is both more dependable and more relevant.

Specifically from the point of view of migration data collection, there are several specific additional enhancements to census taking that are worth considering and that would further add significant *value*, *scope* and *quality* to the population census as the migration data collection instrument of choice for most countries.

Earlier in this section, in the context of approaches to migration data collection through sample surveys, we highlighted two fundamental difficulties associated with probability sampling designs: The first is obtaining an adequate sampling frame. The second is obtaining information on heterogeneity that would enable efficient stratification. Generally, there are no solutions to these issues. However, if a carefully executed methodological approach is followed, then the light census offers a unique opportunity, here. It is a *special implementation* of a *multi-stage census*.

Specifically, one could use this census as what would technically be classified as a sampling frame in a stratified design: One would follow up those respondents who are homogeneous in terms of having experienced at least one migratory move.

In other words, the census is then used in two distinct roles. First, it is used as the *sampling frame*, identifying all members of the population. And, second, it is used as the *source of empirical information to eliminate sample heterogeneity* by subdividing the population into two internally homogeneous groups in respect of migration behaviour, namely those who have experienced one or more migratory moves, and those who have not.

The migratory events that have occurred in the life history of the persons with migratory experience (the *migration stratum*) would *not be further probed* as part of the administering of the basic census questionnaire. Instead, such a person would *merely be flagged in this first round*, and he or she would be *followed up in a dedicated in-depth second round* of special migration behaviour questioning.

This second round would then be under the control of *field staff specially trained for the purpose of collecting retrospective data on migratory moves, their timing and the associated places of residence*.

For this design to be successful, there are two further conditions that must be met.

First, considering that migrants may well be mobile, the follow-up would have to be immediate. That is to say, there should be *no delay between the first and second rounds of questioning*. This requires the use of an effective system of flagging up candidates for the second round. In fact, in countries where the first round is spread over multiple days, the second round of questioning can start immediately where the first round has been completed.

Second, for the reasons discussed earlier in the context of random sampling, the second round would preferably have to include 100% of respondents flagged up in the first round. In other words, in this second round dedicated to capturing migration event histories there should be *no random selection within the migration stratum*.

Random selection within this stratum in such a manner that it leads to acceptable bounds of probabilistic uncertainty would require further more detailed stratification within this migration stratum itself. When the first census round only contains a simple question allowing differentiation between those who do have a migration history and those who do not, then there is no information enabling such further stratification.

Key benefits obtained through migrant targeting and expert questioning typical of this multi-stage approach to census taking are: first, the *opportunity to probe deeper into migration event histories*, beyond merely the single most recent move; and, second, *significantly higher-quality data* in terms of completeness, reliability and precision.

This multi-stage census is an approach that similarly lends itself to the observation of other "difficult" topics in a census. However, many such other topics would, of course, not require that the second stage sample be taken at 100%. Further, if the topic in question is unrelated to migratory behaviour, then the condition of no delay between the first and subsequent rounds might possibly be relaxed. Clearly, however, the longer this delay, then, of course, the more outdated the first round basic census will be as a sampling frame as a consequence of intervening demographic change due to mortality, fertility and migration.

A second approach to add very significant further value to the light census is *census linking and integration*.

Specifically, census linking and integration involves the following approach: Rather than conceiving of each successive census as a self-standing one-off data collection effort, data from successive censuses are linked at the level of individual persons. In other words, census linking and integration effectively generates continuity in the successive data collection processes by *identifying persons uniquely across successive censuses*, and by *properly associating periodically collected data with unique persons*.

It is an approach which, in fact, ties in directly with conceiving the population census itself essentially as a periodic system of population registration. All registered persons are then matched with their values on the socio-economic variables collected from time to time by the national statistical office. Such data can result from periodic censuses, but also from other sources used by NSOs, such as, for example, surveys.

There are various and complementary approaches to realize such census linking and integration in practice.

One logical approach when thinking of the census in terms of population registration, is the use of a nationwide personal identification system. However, even if one instead remains faithful to the traditional interpretation of censuses merely as complete socio-economic surveys, then there are also more technical means to achieve such census linking at the personal level, such as ex-post pattern recognition in the collected data. This is an approach which has seen rapid development in recent years in many diverse areas of application, such as, for example, in signal processing and in digital imaging, in forensic and intelligence research, and in marketing information systems and in data mining.

Successful census linking at 5-year intercensal intervals, for instance, means that retrospective questioning on migration behaviour can be limited to a period no longer than 5 years prior to the enumeration. This reduces the census burden and it places fewer demands on the quality of recollection of respondents, enhancing overall migration data quality.

Census linking and integration is an efficient and cost-effective approach to *collecting cohort life history data*. As such, it clearly is valuable not only for the measurement of migration: To be methodologically rigorous and consistent, all demographic measurement requires data on the occurrence of demographic events in the life histories of cohort members.

More generally, census linking and integration at the level of individual cohort members (persons) results in a *periodic system of longitudinal observation*. Longitudinal observation is fundamental to an in-depth understanding of many socio-economic and other phenomena and processes. The importance of the unique informational value of such data is recognized in many elementary research designs, ranging from, for example, time series analysis in economics and econometrics, to, for instance, longitudinal panel studies in the social sciences, marketing, and medicine and health care.

In combination with a 5-year intercensal interval, census linking and integration provides a bridge that puts the population census conceptually *on as close a par as possible with true continuous observation*, such as is achieved in a continuous population registration system.

However, this is accomplished at a very significantly lower administrative burden and cost.

In addition, continuous population registration systems leave much of the initiative to register with the individual person. This affects in particular the quality of the registration of recent, frequent, and temporary migrants. In the case of a census, on the other hand, the NSO actively reaches out in an effort to enumerate persons.

Further, precisely through the periodic nature of the census there is a built-in mechanism to correct for errors and for obsolescence. For example, through established capture - recapture methods, census linking and integration provides one of the best methods available to assess the quality of the completeness of enumerations. Continuous population registration systems tend to lack a similarly comprehensive systematic error checking mechanism.

Finally, by design, census linking and integration, of course, extends well beyond a continuous population registration system, in that an essential core objective is the integration of observed and recorded persons and their associated socio-economic data. Continuous population registration systems at best hold very limited socio-economic variable data, and where they do, they often lack timely update mechanisms.

In short, by *adding enhanced longitudinal information to the data collected*, census linking and integration adds significant value, both to traditional and to multi-stage censuses. Obviously, this value can be further enhanced by extending the approach to include other data collection efforts as well, such as, for instance, separate sample surveys.

While there are clearly *security of privacy* and *data anonymity* issues involved in census linking and integration that need adequate addressing, such issues are not different in principle from those already encountered in traditional census taking.

Let us briefly summarize. As we have seen, in most countries the population census will be the primary source of data on internal and international migration. However, we have also seen that traditional approaches to census taking come with a range of limitations in this regard.

Through the combination of a set of complementary and mutually reinforcing approaches, including the *light-weight core census* repeated at *5-year intercensal intervals*, the special implementation of the *multi-stage census*, and *census linking and integration*, such limitations can be largely overcome or at least be significantly mitigated. Moreover, as discussed, these three approaches yield major additional operational and informational benefits. Both individually and together, they can considerably enhance the importance of the census as the source of choice for comprehensive data on ongoing and historical migration processes.

6 Census Data Processing and Migration

The collection of statistical data is a fundamental cornerstone of well-informed and transparent evidence-based policy making, planning, management, monitoring, evaluation and assessment. Developing an effective evidence base involves more than simply collecting data: The data gathered must also be *relevant* to the task in hand: They must be able *adequately to inform* the discussion, debate and decision making *as and when the need arises*.

In other words, relevance has two aspects. First, there is the matter of identifying data that are of *substantive* importance. However, second, relevance also has a *procedural* side: The collected data must be organized and managed in such a way as to *facilitate* the development of the necessary insights on a *demand driven* basis. Thus, a national statistical office (NSO) has a dual role, one as the organization responsible for *gathering official statistics*, and one as a *service provider delivering* information that *meets the needs of the users*.

> Let us give an example. Let us, for instance, assume that the debate is about the optimal management and development of a major slum settlement. One of the first questions likely to arise will be the elementary matter of the slum's population size, composition, distribution and dynamics. A key issue then, of course, is the geographic mobility *into*, *within*, and *out of* the settlement in question. As regards data on geographic mobility, it will be clear from section 5, above, that an obvious key source is the census of population.
>
> Now, let us assume that the national statistical office has followed good practice by adhering to the United Nations *Principles and Recommendations for Population and Housing Censuses* (see section 8, below, for more details). As far as the collection of migration data is concerned, these recommendations

allow some latitude, and, at best, data will have been collected on the last move only. In addition, all recorded places of previous residence will have been pre-classified in major civil divisions.

Clearly, however, data on the most recent move only do not allow one to adequately inform the debate on the nature, characteristics and dynamics of the inward, internal and outward migratory behaviour and trajectories by slum dwellers experienced over recent years leading up to the present time.

Further, the pre-classification of all places of origin of migrants by major civil division will severely restrict the ability to study internal slum mobility in adequate geographic detail. In fact, if the slum settlement happens to be located entirely within one single major civil division, then there will not even be any record at all of the internal mobility within the slum.

Next, if the most recent population census was conducted several years ago, then inevitably there is a total void of information on any more recent developments after the census was taken. The lack of availability of any recent information is an issue that, on average, will be the more severe, the longer the NSO's intercensal intervals.

Finally, major practical problems can arise in actually delivering the necessary information, unless the NSO is properly geared up to producing non-routine tailor-made data sets meeting the specific information demands arising from the slum debate in question.

In respect of the migration flows affecting the slum settlement, such demands might involve, for example, a combination of purpose-defined migration-defining areas (MDAs), specifically selected cohorts, specifically selected time frames, specifically selected trajectories, possibly with special filtering of temporary moves, of frequent movers, of moves conditional on earlier migration behaviour or conditional on other attributes, and so on.

In day-to-day practice, many NSOs find it difficult to meet demands for the assembly of such non-routine tailor-made data sets in a manner that is not only dependable, but also efficient and timely, as well as cost effective. Yet, such demands are inevitably associated with a realistic and in-depth understanding of historical and ongoing migration processes shaping the slum settlement in question.

Clearly, therefore, as a consequence of such *substantive* and *procedural* limitations and bottlenecks, overall the *relevance* of the collected data can in practice be severely limited.

Let us explore these issues in some more detail.

In the previous section 5 we have outlined a number of operational avenues directed at enriching the census as a tool to inform societies in the area of migration -- and at the same time in many other areas: We described a complimentary set of approaches, combining more imaginative and innovative census data collection methods leading to enhanced continuity, scope, depth and quality of the resulting data on ongoing migration processes. The approaches entailed both substantive and procedural improvements.

However, these improvements in themselves do not guarantee improved relevance. Once collected, the data must be processed to produce relevant information. Until recently, census processing was predominantly *supply-driven*. It involved the production of a large number of pre-defined printed standard census tabulations for the country and for the provinces or other major administrative regional units. Often, these standard tabulations were supplemented by one of more descriptive and interpretative topical reports. These outputs were deemed useful more or less by definition. All too often they served as the intrinsic justification of the census operation, as its primary raison d'être.

It is a line of thinking that has its origins in the days where NSOs had no IT systems, and where the processing of the completed census questionnaires was done manually. In practice, this prohibited a re-run of the raw data to produce alternative data sets on special request at a later date.

However, with the introduction of non-paper-based storage systems, originally in the form of magnetic tapes, there was no longer any justification to limit the census output to a pre-defined set of printed standard tabulations. Yet, the legacy approach of the 1950s and earlier has proven remarkably persistent, especially in the developing world. It is, in fact, still clearly in evidence in the Principles and Recommendations of the United Nations, even for the 2010 worldwide round of population and housing censuses (see section 8, below).

Obviously and importantly, though, in today's rapidly changing socio-economic conditions, such a limited and supply-driven approach can no longer have a rightful place. Proper evidence-based policy making, planning, management, monitoring and assessment poses challenges and information demands which cannot be comprehensively anticipated at 10-year intercensal intervals. In order to meet such needs for substantiating empirical information, NSOs have to adopt a *demand-driven*, *facilitating* and *service-oriented* position.

From the perspective of information relevance, the problems with the traditional supply-driven attitude in census processing are twofold. First, *rarely does such a pre-determined standard publication programme address any specific needs*, such as, for instance, the needs outlined at the start at this section in the example of slum management and development.

In fact, the resulting perceived lack of sufficient purpose and usefulness in meeting real-world information demands, especially when set against the high costs involved, explains at least in part the low priority awarded to the taking of periodic population and housing censuses in too many developing countries.

Second, the common traditional 10-year intercensal interval, not seldom combined with a long stretched-out data processing programme, means that *the resulting information can rarely be described as timely*. If recourse is taken to census information, as will usually be the case in the study of ongoing and historical migration processes, then inevitably one often deals with dated or simply outdated information. Clearly, as discussed earlier, timeliness is an essential element of relevance.

In section 5, we already outlined a range of substantive and procedural advantages of reducing the intercensal interval, and we discussed operational approaches to making such a reduction practically feasible.

That therefore still leaves us with the issue of the ability of NSOs *cost-effectively, timely and efficiently* to *answer tailor made requests for information*.

Fulfilling such requests requires the flexibility to meet client requirements *on demand* and *as a matter of routine*. And, equally importantly, it requires the ability to meet such requirements *dependably* and *transparently*, *guaranteed precisely to satisfy the required data specifications*.

The ability of an NSO to provide such services is an issue that is probably nowhere more pressing than in the case of migration. Information on ongoing and historical migration processes which adequately addresses real needs, will often necessitate the *compilation of data sets* that require *very considerable and relatively complex processing of the raw census data*.

For instance, it will usually require the compilation of specific demand-driven definitions of *migration-defining areas* (MDAs); the compilation of *specific cohorts* further characterized by a selection of relevant attributes and perhaps defined by place of residence not at the time of enumeration; the *definition and selection of the migratory moves* that are relevant for the issue at hand -- maybe specific trajectories only, perhaps on a given time interval only, possibly only n^{th}-order moves, maybe only multiple moves occurring with a specified frequency, perhaps only moves conditional on specified earlier migratory or other behaviour; and so on. Census linking and integration, discussed in the previous section, with a view to obtaining better quality and more informative longitudinal data structures, adds yet further to the possibilities and the associated data processing challenges faced here.

To produce such information efficiently and with the confidence that it has been compiled accurately from the raw collected data, puts very considerable demands on *information systems design* and on *information retrieval procedures*. At the same time the issue here ties in directly with the availability of *adequate IT skills* at national statistical offices.

When set against the costs and efforts involved, for instance, in taking a full population and housing census, then realizing a proper and an adequately staffed integrated information systems infrastructure and facilities may seem trivial. All too often still, however, especially in the developing world, the matter is treated as of secondary importance. Yet, as will be clear from our discussion, the relevance of the expensive data collection effort crucially depends on this as a *necessary condition*.

There are many aspects that should be considered when designing and implementing an adequate information systems infrastructure. From our perspective, the compilation of on-demand tailor-made data sets for the measurement and analysis of migration processes, there is one issue that is of particular and direct importance:

Many NSOs have opted to use tailor-made database systems for the entry, processing, retrieval and tabulation of their census data. A well-known example is CSPro, an application promoted amongst others by the United States Census Bureau and made available at no cost. Familiar older but similar application packages are IMPS and ISSA.

The underlying proprietary database structure of these applications is based on household and person records typical of traditional simple card files. In the IT industry, for serious information systems such database structures have long been found cumbersome to handle and prone to data processing and retrieval errors. They are obsolete and they have been superseded by *logically structured relational database designs*, accompanied by *powerful and efficient set-theory based query languages* such as SQL.

Industry-standard relational database designs and industry-standard database query languages routinely allow the dependable organization and accurate retrieval of complex tailor-made data sets from a census database, such as the requests typically associated with the measurement of migration.

Further, given a well-structured and properly documented census database, tasks such as this can be completed in a matter of hours at the most, including the formulation of the query itself.

In addition, census linking and integration discussed in section 5 is in practice a realistic option only when exploiting the expandability and transparency of the logical modular structures inherent in a relational database design. This equally applies to the practical implementation of multi-stage censuses as described in section 5, and also to the linking and integration of censuses with yet other data sources.

On the other hand, meeting even the most basic of requests for special on-demand tailor-made data sets with outdated database systems such as CSPro proves highly cumbersome, time consuming and error prone. Experience has shown that in practice it can easily take weeks of special software development, testing and debugging to retrieve, organize and produce the type of data sets that are typically required for the measurement of migration processes.

More generally, the adoption of industry-standard systems and approaches to information system design relegates to the past the practical inability -- on the usual time and cost grounds -- to produce relevant information specified on demand from a census dependably and efficiently.

However, there is another major reason to adopt industry-standard approaches to information systems design and information retrieval. Traditional proprietary database designs make heavy demands on *dedicated specialist in-house IT skills* at national statistical offices throughout the useful life time of the collected census data.

First, this is costly because of the ongoing necessity to maintain dedicated application-specific expertise and because of the need to provide associated specialist staff training. Second, experience teaches that a national statistical office must necessarily assume the inevitability of the fact that it is structurally unable to retain its skilled IT staff in view of the higher wage levels typically offered fur such skills in the private sector.

Relational database design and query languages such as SQL are simple to learn and they are core instruments in the standard tool kit of every decent first-year undergraduate IT programme. Consequently, *the necessary skills are abundantly available* and *easy to acquire*, even when the disposable financial budgets are severely limited. And in the case of low IT staff retention rates, the skills are *easily replaceable*.

It is important to note that a *<u>relational database design</u> and the <u>query language</u>* are *<u>separate from</u> and <u>independent of</u>* the -- usually proprietary -- *<u>software application</u>* that is used to implement the design and to make the queries. A proper structural database design can be implemented in any number of such different modern software applications.

Here we therefore have a fundamental difference with obsolete packages such as

CSPro, where database design and software implementation are intertwined, ultimately resulting in major negative cost and skills implications in the long term. In addition, fully competent commercial application software for the implementation of modern relational database designs is now available at a cost which is insignificant relative to the costs associated with the collection of the census data.

Further, in order to ensure *independence* between the database itself and the software application selected for its implementation, and in order to ensure the *integrity* of the costly, invaluable and irreplaceable census data, the four basic IT tenets remain elementary:

(1) proper *clear*, *complete* and *up-to-date documentation* of the *functional* and *technical database design specification*;
(2) *routine data back-up* in a *software application independent format*; this is commonly achieved through simple data export in industry-standard ASCII-encoded CSV (comma-separated values or comma-separated variables) format;
(3) maintaining copies of these back-ups onto a *carrier medium whose integrity and currency* (functionality and hard-ware compatibility) *is ensured* from time to time; and
(4) *safe-keeping* of the back-ups in a *secure location* which, in addition, is *away from the premises where the originals are maintained*.

However simple and obvious these rules may be, in practice they are more often violated than not. NSOs should therefore adopt and enforce rigorous and unambiguous policies here.

Finally, a similar strict attitude should be taken in respect of *data protection against any unauthorized access*, and in particular as regards the *safeguarding of the privacy of the respondents* who have supplied their personal data. Here, two developments especially pose ever increasing risks.

First, this is the increasing network interactivity with and between information systems. The second risk is that associated with the ubiquitous use of increasingly compact high-capacity portable mass storage devices.

7 Measurement and Analysis of Migration -- An Overview

The primary *key target* of the measurement of migration processes is to obtain *reliable information on* *migration intensities* (*instantaneous migration rates*) and on *flows and stocks of migrants in absolute numbers* *as time progresses* and *specific by geographic trajectory*. All other information on migration processes (see section 1) derives from or builds on this foundation.

The basic measurement procedure of migration processes is *demographically general*: Its principles apply similarly to, for instance, the measurement of mortality or of fertility.

There is, in fact, a direct relationship with some of the key concepts underlying life table construction in the measurement of mortality. However, several mathematical and methodological inconsistencies and weaknesses, as well as unnecessary and avoidable approximations, have crept into today's common approaches to measuring mortality. The measurement procedure as described in this section, in contrast, has been developed to be *mathematically consistent*, *methodologically rigorous*, and *empirically maximally informative*.

In this penultimate section of chapter 2, we shall briefly outline some of the key elementary principles in a step-by-step fashion. The aim is to provide a bird's eye overview, supporting the material on data that represents the core of this chapter. In order to facilitate accessibility to as broad a readership as possible, the present section avoids any mathematical or complex technical operational detail. As a brief and non-technical summary overview it complements chapter 1 where these matters are treated more in depth.

In line with the earlier parts of this chapter, there will be some emphasis in this section on the use of data that derive from population censuses. Further, the aim is to keep this summary presentation at an introductory level. Therefore, we shall restrict our attention here mainly to data pertaining to the most recent move (last move data), paying no explicit attention to higher order (2^{nd}, 3^{rd}, …) moves. However, the extension to such other moves, and dealing with multiple moves more generally, is, in fact, straightforward and proceeds according to the same methodological and operational principles. Xu-Doeve (2006), in particular, provides more details on this issue.

We begin the measurement procedure as follows: At some chosen point in time, called the *benchmark time point*, we assemble the persons who are then *resident in a given place* and who have *similar or identical attributes*, the most important among which is *age*, in internally homogeneous *cohorts*.

The *benchmark point in time* is usually simply denoted by $t = 0$, and in the measurement procedure time is then conveniently reckoned from zero. In the case of population census data, it is often practical, but by no means necessary, to choose the point in time of the census enumeration as this benchmark time point.

It follows from the above definition of a cohort that choosing another benchmark time point has a direct bearing on the composition of the cohorts. This fact can usefully be exploited as an analytical device.

The selected *place of residence* is operationally defined in terms of a *migration-defining area* (MDA). The specific delineation of MDAs is determined by the scope and purpose of the analysis that is to be carried out.

As to the selection of the *attributes* of persons that should be taken into account in the assembly of persons into internally homogeneous cohorts, here, too, the choice is determined by those characteristics that are considered relevant in the light of the objectives of the analysis in question.

As noted, the primary attribute to be considered is age: For both methodological and empirical reasons, assembling persons into groups that are homogeneous with respect to age, that is, classifying persons into age groups, should receive a high priority.

Examples of further attributes by which persons may be grouped might, for instance, include: the same sex, similar earlier migration behaviour, the same place or country at birth, the same nationality / citizenship, the same residence permit status, the same or similar educational background, similar economic activity, similar income, similar motives for migrating, similar intentions regarding staying permanently or not, and so on and so forth.

The *key to measurement* now is the analysis of the occurrence of the set of, or, more commonly in applied practice, a selective subset of the individual *migratory events* (*migratory moves*) in the life histories of individual cohort members as time and age progress.

In other words, the essence of the measurement of migration processes centres on establishing the *timing*, the *MDA of origin* and the *MDA of destination* of the *migratory events* experienced by cohort members *as time progresses* and *as the cohort ages*.

In the case of population censuses, the necessary data are obtained through *retrospective questioning* on *places of previous residence*, and on associated *durations of residence* or, alternatively, on associated *dates of moves*.

In practice, the subset of events recorded in population censuses usually comprises only the most recent move of cohort members. Empirically, this is highly restrictive. It is an issue which is discussed in section 3, above.

Operationally, the measurement procedure centres on *observing* and *formally capturing* the *attrition* of the cohort *over time and age*, *specific by migration trajectory* (geographical route or origin-to-destination path).

Attrition is defined here as the *decline in the number of cohort members* who have *not or not yet experienced the migratory event* under study *as time and age progress*.

In charting and interpreting attrition, we can distinguish between forward analysis and backward analysis. In *forward analysis*, such attrition is caused by out- or emigration. In *backwards analysis* we apply the device of *time reversal*, and thus also of migration process reversal, including the reversal of origin and destination.

Time reversal involves changing the order of time, that is, letting the clock run backwards from the benchmark time point. Time reversal in the analysis of attrition compares to playing a film of the migration process backwards: A person actually entering an MDA is perceived, and mathematically described, as exiting. In backwards analysis, attrition, therefore, reflects in- or immigration which is merely mathematically formulated, and therefore measured technically, as if it were out-or emigration *as time is traced backwards*.

Backwards analysis is generally the most practical approach in the case of *retrospective data* on past migration behaviour, such as the data typically collected through questioning in censuses and surveys. Forward analysis is most convenient when event data have been recorded *in real time*, such as in a continuous population registration system.

Mathematically, *attrition* is described by what is commonly, but restrictively, known as a *cohort survival function*. In the case of sufficiently internally homogeneous cohorts, such a survival function itself can be shown to be purely a function of a cohort's trajectory-specific *migration intensity*, that is, of its *trajectory-specific* *time-continuous* ***instantaneous migration rates** function*.

Such a time-continuous instantaneous migration rates function is commonly denoted by $\mu(t)$. Since age runs in parallel with time, these functions are not only continuous functions of time t, but also of exact age x.

We re-emphasize that such functions $\mu(t)$ are *specific* not only *by cohort*, but also *by MDA to MDA* (origin to destination) *flow*. Clearly, furthermore, if a cohort has been defined specific by any further attributes, such as age, sex, and/or any other attributes, then these instantaneous migration rates are similarly specific.

An instantaneous rates function $\mu(t)$ is more general, more valid, and empirically more informative than the empirical annual rates M_x which we frequently encounter in standard demographic textbooks and which are common in applied demographic analysis. Such traditional rates of the type M_x are defined as constant in a given year and as constant on a given age interval. Usually, in addition, they are evaluated in ways that are not conceptually and mathematically consistent and which, moreover, incorporate poor approximations. Instantaneous rates functions of the type $\mu(t)$ with which we are dealing here, however, do not suffer from any such drawbacks. This matter is discussed in detail in chapter 1.

> We note that *attrition*, *cohort survival functions* and *instantaneous rates functions* (*process intensity functions*) all three are general concepts. Here we in fact have an immediate link with, for instance, mortality: In the case of mortality we similarly encounter a cohort survival function in classical life table analysis, namely in the mathematical formulation of $\ell(x)$, the survivors from birth. This function is similarly governed by a corresponding instantaneous mortality rates function (mortality intensity function). Obviously, the specific shape of an instantaneous rates function characteristic of cohort mortality behaviour (the mortality intensity function) will be quite different from an instantaneous rates function typical of cohort migration behaviour (a migration intensity function). And, of course, the same therefore applies to survival functions characterizing cohort attrition due to mortality and migration, respectively.

Now, *given a set of empirical cohort migration attrition data* -- by definition specific by MDA to MDA trajectory -- as time and age progress, then it is easily shown mathematically that this *instantaneous migration rates function $\mu(t)$* is *simply and directly obtained* by a procedure well known from statistics as "***curve fitting***".

In the case of data from a full enumeration, the recommended procedure is least-squares estimation. If, on the other hand, the data are based on random sampling (probability sampling), then maximum likelihood estimation is the recommended technique.

In applied practice, cohort attrition data sets may reflect not only attrition due exclusively to migration, but also due to the *simultaneously operating competing force* (or *risk*) *of mortality*. This will, for instance, always be the case when we have retrospective data on past migration behaviour from population censuses and surveys: Clearly, if we have data obtained through retrospective questioning on timing and trajectory of past migratory events experienced by cohort members, then, of course, these data pertain only to those persons who have survived up to the point in time of the enumeration. Consequently, the resulting attrition data will represent the combined effects of both the force of migration and the force of mortality acting as time and age progressed.

There are two methods to *eliminate the effect of this competing force of mortality* on the measurement of migration intensities. Both require prior knowledge of cohort mortality, preferably from a cohort life table.

The first is the well-known demographic method of *reverse survival*, undoing the effect of mortality on the observed cohort attrition data. Reverse survival is therefore applied *before curve fitting*.

The second method, instead, *immediately proceeds with curve fitting* first. The result is a compound instantaneous rates function reflecting the simultaneous operation of the forces of mortality and migration. The *instantaneous mortality rates function* is then *subtracted* from this compound instantaneous rates function obtained through the curve fitting procedure. Since intensity functions reflecting competing demographic risks are additive, the resulting difference is the desired instantaneous migration rates function.

We note here that this second method is, in fact, an application of a principle of broader general importance. Recall, for instance, that for any given cohort each origin-to-destination trajectory is characterized by a trajectory-specific force of migration, or, equivalently, by a trajectory-specific instantaneous migration rates function $\mu(t)$. Consequently, this additive property of demographic forces is also a core instrument in the measurement and analysis of the distribution of a cohort over the set of MDAs of destination as time and age progress.

It will be clear that this approach to the measurement of migration rates is *general*. It applies equally to the measurement of other demographic rates, such as, for instance, *mortality rates* and *fertility rates*.

Normally, the case of mortality is the simplest. Usually one does not have to reckon with the occurrence of any competing events, here. In addition, the event of death can occur only once in the life history of a person. As a consequence, one does not encounter the equivalent of multiple moves or of multiple births in the life history of a person that can occur in the study of migration and fertility, respectively.

The approach also extends to many other disciplines, as well, such as economics, engineering and logistics, where many *formally similar hazard rates* are encountered.

Further, the approach also provides consistent and complete answers to several common questions in demography, such as, for example, the issue of what has become known as the "tempo" component and the "quantum" component of observed demographic change.

This is a direct result of the underlying *event-based perspective* and the associated rigorous *cohort orientation*.

Quantified instantaneous migration rates functions $\mu(t)$ are significant themselves, namely, in terms of the comprehensive detailed information they provide characterizing the *intensities* of migration processes and the *dynamics* in these intensities *over time and age*. However, these functions also hold the key to obtaining full information on migrant flow and stock numbers:

It can easily be shown mathematically that, given this function $\mu(t)$ describing the instantaneous migration rates over time and age, then the ***absolute numbers of migrants***, *specific by cohort, by trajectory, and by any other attributes considered*, that is, migrant ***flows*** and ***stocks***, follow direct and also *as continuous functions of time and age*.

Thus, we also obtain complete information on the *dynamics* over time and age of the migration process in terms of absolute numbers of persons involved.

Data on *recent migrants* are often highly and disproportionally incomplete, and the measurement procedure can also be used to *adjust for such incompleteness* in a straightforward and transparent manner. The method to do so is explained in chapter 1.

We note that the adjustment for the incompleteness of observed recent migrants is *possible only* when *events* (migratory moves) have been recorded. If one instead chooses to record and count *migrants* (*persons*), defined by an attribute used as a proxy indicative of migration behaviour as discussed in section 4, above, then it is *impossible to estimate this incompleteness and to adjust the data for it*.

The measurement of migration as outlined in this section is discussed in detail in chapter 1. A full in-depth treatment, including a modern account of the underlying mathematics, can be found in Xu-Doeve (2006). Both also contain detailed empirical examples using data on last moves (duration of residence and place of previous residence) from Thailand's 1970 population and housing census.

To conclude this section, let us briefly summarize the key results in two main points:

First, following the approach to data specification, to data collection and processing, and to demographic measurement outlined in this chapter, the migration processes under study can be *empirically* ***completely*** *captured, identified and characterized*.

Second, the recovery of this information is achieved in a manner that is *fully transparent*; *conceptually, mathematically and demographically consistent*; and *methodologically rigorous*.

8 References and Selected Annotated Key Related Materials

The present chapter has been prepared with readers with a limited background and/or interest in strict formal demographic analysis foremost in mind. A basic familiarity with the traditional demographic paradigm in respect of data and methods is recommended. Useful textbooks are, for example:

> Shryock, H S, J S Siegel and associates (1971) *The Methods and Materials of Demography. 2 Volumes*. US Bureau of the Census
>
> Siegel, J S and D A Swanson, eds (2004) *The Methods and Materials of Demography. 2nd Edition*. Elsevier/Academic Press

In the context of the present chapter, these two textbooks are useful as background material for two reasons in particular. They pay considerable attention to data, and they each contain chapters on internal and international migration. We note that these chapters on migration reflect a traditional approach to measurement and analysis. However, as such, they provide a valuable contrast to the modern methods that are the subject of the present chapter.

The present chapter itself centres on *migration data* and related issues. It complements chapter 1 which outlines the theoretical, methodological and operational aspects of modern *methods of measuring* internal and international migration.

We note that chapter 1 was presented earlier as Xu-Doeve (2007). A condensed version of that chapter, Xu-Doeve (2008), was written at the request of the Statistics Division of the United Nations.

Together, chapters 1 and 2 provide a comprehensive introductory overview of modern methods of measuring migration. For in-depth further reading, the reader is referred to Xu-Doeve (2006). This book also gives a modern and complete account of the underlying mathematical theory. In addition, it contains extensive references to methodological and empirical literature materials reflecting common current practice.

Xu-Doeve, W L J (2006) *Methods of Measuring Internal and International Migration*. ANRC Publishing. ISBN-13: 978-90-8802-001-8

Xu-Doeve, W L J (2007) *The Basic Principles of the Measurement of Migration Using Population Censuses*. Paper presented at the International Conference on Migration and Development (the Fifth Valentey Lecture, Lomonosov Moscow State University), Moscow, 13-15 September 2007. Forthcoming with minor editorial changes in the demographic journal *Genus*

Xu-Doeve, W L J (2008) *The Applied Measurement of Migration*. Paper invited by the United Nations Statistics Division (UNSD) and presented at the 56th Session of the International Statistical Institute (ISI), Lisbon, 22-29 August 2007. Revised 2008. Published in the Proceedings of the 56th Session of the International Statistical Institute

Next, we list the *principal international recommendations* generally used as the primary standard source of reference by national statistical offices when deciding on how to collect data on internal and international migration.

United Nations (1970) *Manual VI. Methods of Measuring Internal Migration*. United Nations

United Nations (1997) *Principles and Recommendations for Population and Housing Censuses. Revision 1*. United Nations

United Nations (1998) *Recommendations on Statistics of International Migration. Revision 1*. United Nations

United Nations (2007) *Principles and Recommendations for Population and*

Housing Censuses. Revision 2. United Nations

United Nations Economic Commission for Europe (2006) *Conference of European Statisticians Recommendations for the 2010 Censuses of Population and Housing*. Prepared in cooperation with the Statistical Office of the European Communities (EUROSTAT). United Nations

United Nations Economic Commission for Europe and Statistical Office of the European Communities (1998) *Recommendations for the 2000 Censuses of Population and Housing in the ECE Region*. United Nations

We note that, at the time of writing, United Nations (1998) is in the process of being revised. This revised edition is due for publication in the second quarter of 2008.

There are important methodological differences between these internationally recommended methods and practices, on the one hand, and the principles and guidelines in the present book and in Xu-Doeve (2006), on the other. It is therefore useful to compare and contrast these different approaches.

For example, one such detailed comparison, referring to United Nations (1997) and United Nations (1998), is available in Xu-Doeve (2006). We note here that United Nations (1997) has now, of course, been superseded by the second revision, United Nations (2007). Important issues remain, however: In terms of the measurement of migration processes, apart from less pressing matters, there are at least three major and fundamental concerns.

First, as regards the interpretation of the concept of a place of usual residence, the notion of a minimum stay of 12 months is maintained in the international recommendations of the United Nations (1998, 2007). This issue is discussed in section 3 of this chapter.

Second, as to internal migration, in the international recommendations of the United Nations (1998, 2007) questioning on durations of residence and associated places of previous residence is not clearly preferred over questioning on place of residence at some fixed date in the past. We discussed this topic in section 4, above.

In particular in the United Nations (1998) recommendations, finally, the measurement of international migration is narrowly interpreted as no more than a count of alien or foreign persons. This issue, too, is discussed in section 4 of the present chapter.

Such methodological issues result in *data sets* that *can severely limit the ability* *comprehensively to measure, analyse and interpret actual ongoing or historical processes of migration*.

It is also interesting to compare and contrast the recommendations for the 2000 and 2010 population and housing censuses in terms of their respective approaches to the collection of data on internal and international migration.

For example, during the preparations of the recommendations for 2010 for the UN-ECE region, there have been detailed exchanges and discussions on the matter with UN-ECE, EUROSTAT, and members of the Conference of European Statisticians, building on the methods and materials now available in Xu-Doeve (2006). The 2010 UN-ECE recommendations, compared to the recommendations for the 2000 census round in this region, reflect this in a welcome, albeit incomplete, shift in emphasis from traditional person-based counting to modern event-based observation.

From a methodological point of view, the following *flagship OECD publication on international migration* is also worth exploring.

> Organisation for Economic Co-operation and Development (2007) *International Migration Outlook. Annual Report 2007 Edition (Sopemi 2007 Edition)*. OECD

This is an annual publication, and prior to 2006 it was published under the title "*Trends in International Migration*". These thorough in-depth reports are compiled and edited by OECD, and an important input to this work are the contributions by a dedicated network of local correspondents in OECD member states.

The reports highlight some of the constraints which the data that are currently available in OECD member states place on measurement and analysis, especially when these data are used in an internationally comparative perspective.

For further in-depth background details on data issues and on the associated methodological aspects and developments that underlie these reports, the following OECD papers are also useful. They are available from the authors at OECD.

> Dumont, J-C and G Lemaître (2005) *Counting Immigrants and Expatriates in OECD Countries: A New Perspective*. OECD Social, Employment and

Migration Working Papers No 25. OECD

Lemaître, G (2005) *The Comparability of International Migration Statistics. Problems and Prospects*. OECD Statistics Brief No 9. OECD

Lemaître, G, T Liebig, C Thoreau and P Fron (2007) *Standardised Statistics on Immigrant Inflows. Results, Sources and Methods*. Informal Paper. OECD

Lemaître, G and C Thoreau (2006) *Estimating the Foreign-Born Population on a Current Basis*. Informal Paper. OECD

Although the title might suggest otherwise, the scope of these annual *Trends / Outlook* reports is not truly international in the sense that their focus is more limited in two important respects: They centre on *immigration* into *OECD member countries*. OECD membership is made up of a group of currently 30 countries. Most of these are high-income countries, although there are some notable exceptions, such as Mexico, South Korea, Turkey, and four of the former socialist East-European states.

In several methodological respects, the approach taken by OECD differs fundamentally from the one outlined in the present chapter. However, there is also an interesting recent development in the approach taken by the OECD Secretariat to measurement that holds at least the potential of convergence. We shall explore these issues in some depth here, because they usefully complement the material presented in the earlier sections of the present chapter.

Importantly, first and foremost there is no disguising an overriding interest in the *Trends / Outlook* reports in *resident alien foreigners* present within the borders of the OECD member states. We refer to section 4, above, for a discussion of this interpretation of international migration.

Unsurprisingly, therefore, the familiar traditional approach to the measurement of international migration, the direct count of persons, is well represented in the reports. In this approach, the *migratory behaviour* experienced by persons, that is, the migratory events or migratory moves as time and age progress, is *not itself observed*. Instead, personal characteristics (*attributes*) are observed which are taken as *proxies of migration behaviour*: In order to discriminate between persons who are characterized as migrants and those who are not, persons are defined as migrants on the basis of their attribute country of birth (foreign-born residents) or on the basis of their attribute nationality / citizenship (foreign residents or resident expatriates).

However, since the *2006 Outlook* edition there is also a new development in the choice of data sources and of data types: In addition to their traditional sources, the *Outlook* authors have started to make use of data based on *permanent residence permits* granted by OECD member countries.

According to OECD, this partial shift in data preference has been motivated by two considerations. First, there is the desire more prominently and convincingly to highlight permanent immigration: The authors wish to be able to focus on permanently resident foreign immigrants, as opposed to temporary ones, among the OECD member country populations. Second, the authors wish to have such data classified by specific category under which entry permission has been obtained, such as work-related, family-related, humanitarian, and so on.

The *Outlook* authors feel that they are unable adequately to distil this information from the traditional data sources that are available to them, including population censuses, population registration systems, and surveys. In United Nations (1998) there is a recommendation that the distinction between short-term and long-term migrants be made on the basis of a duration of residence criterion. Specifically, the recommendation is that a migrant be defined as a long-term migrant if the duration of residence is at least one year. (Of course, the specific choice of one year as the cut-off point is entirely arbitrary.) Despite this international recommended standard, OECD member countries differ in their own definitions of what constitutes a long-term migrant. Consequently, the resulting data on long-term migration are not comparable across OECD nations. In addition, information on the associated reasons for migrating can also be impossible to obtain from the traditional data sources available to OECD. Indefinite residence permits and residence permits which in principle are indefinitely renewable, on the other hand, have an implicit built-in permanency criterion, and, furthermore, they are issued for clear-cut reasons, so the *Outlook* authors argue.

Of course, in its desire to obtain internationally harmonized and thus comparable statistics on long-term foreign migration, the OECD in many ways simply shifts the problem. National regulatory systems and practice as regards the grant of residence permits vary enormously across OECD member countries. Let us give just one example to illustrate this.

> The right to permanent residence, for instance, in itself is not a standardized concept. For example, in most cases European nations have policies to grant temporary residence first, irrespective of the ultimate intentions of the authorities or of the applicant. Subject to conditions and after a specified time period, this permission may subsequently be extended by one or more further terms or on an indefinite basis. Actual long-term residence then only becomes apparent and measurable after such a status change. In quite similar cases, the traditional settlement countries of the New World, on the other hand, may grant the right of permanent residence immediately on entry. However, within each of these two broad groups of countries, the Old and the New World, actual

policies and applied practice show very significant further variation, both in terms of principles and in detail.

As to the concept of long-term residence, it is worth noting here that the *Outlook* authors view this in a very strict sense, namely as permanent immigrants who become members of the resident population. On this basis, they choose to exclude temporary but very long-term residents such as, for example, students on academic degree programmes. For several OECD member countries, as a consequence, the resulting OECD statistics can be very significantly different from published national official statistics on long-term international migration.

In addition, as a means to measure migration, residence permit data suffer from many well-known further drawbacks and deficiencies.

For example, the date of the grant of permission to take up residence does not necessarily coincide with the date of the actual move. Both national regulations and practices and personal choice of the migrant can be a cause. Usually it is the date of granting permission which ultimately shows up in the statistics. In fact, permits issued may even not be taken up at all. Sometimes administrative backlogs can also play a role. When focusing exclusively on long-term permits, the matter of the true date of immigration is further compounded by renewals and by status changes from temporary to indefinite and permanent permission to stay.

Next, not every foreign settler in a country always needs a residence permit. Permit systems apply only to regulated inflows, that is, to inflows subject to regulatory control. For example, in the case of flows that occur under free movement regimes, residence permits may not at all, not necessarily, or not systematically be issued.

Importantly, further, residence permits only cover duly processed immigrants. That is to say, they only cover immigrants who are observed administratively and who are officially authorized to immigrate. Any unauthorized (illegal, irregular or undocumented) migration is never captured.

In addition, since outflows are normally unregulated, residence permit data can usually tell little if anything about foreigners leaving the country. In the case of residence permit data, this lack of information cannot be compensated for in any systematic manner by resorting to corresponding residence permit based immigration statistics of the receiving countries: Often, a departing foreigner will be leaving for his or her home country where no residence permit restrictions will apply to this person. In such cases, the departing foreigner will, therefore, not show up in the corresponding residence permit based immigration statistics of the country of arrival. So, in the absence of adequate data on departures from residence permit sources, over time the resulting cumulative statistics on foreign additions to the resident population will differ more and more from the actual numbers of foreign residents.

> Also, statistically it can prove difficult in practice unambiguously to identify and match persons and permits issued. For example, a person may have been granted more than a single permit on different grounds, such as employment and family reunification, or there may have been other status changes that hinder accurate statistical identification and reconciliation. In addition, the occurrence of a migratory move is a potentially recurring event in the life history of a person. Consequently, a person may be issued with a residence permit on the same grounds multiple times. In the case of seasonal and other temporary work, for instance, this can be a common cause for discrepancies between the numbers of permits granted and the numbers of actual migrants.
>
> Yet a further difficulty is that the grounds on which permits are issued do not necessarily agree with actual real use. Legally, permits may already allow multiple use. For instance, a permit issued on the grounds of family formation may sometimes automatically extend the right to work. Also, subsequent use of a permit may not agree with the original grounds, legally or otherwise. In the previous example of family formation, the family relationship might end while the work continues. In fact, actual real use may well completely deviate from the grounds on which the application was made in the first place. For instance, permits for reason of study are often easier and cheaper to obtain than permits for work and settlement, and abuse of this is not uncommon in countries of origin such as China.

Clearly, the use of residence permit data for the measurement of international migration is a statistical minefield. This is a situation that is not uncommon when using *administrative data in lieu of dedicated statistics*. Usually, such administrative data are not primarily collected for statistical purposes. In addition, contrary to, for instance, population census data, their collection takes place by a third party and outside the direct conceptual and operational control of national statistical offices or other research institutions. In the case of residence permit data, there is the specific further matter that in many countries the question of permission to immigrate has become a politically sensitive issue.

As a consequence, obtaining results from residence permit sources that are meaningful and unambiguous, complete and correct, and internationally harmonized and comparable, is a matter which requires an intimate familiarity with underlying national regulatory principles and with actual applied practice. This is made all the more difficult, since national regulatory systems and practice tend to be the subject of frequent major and minor changes and adjustments. In addition, actual applied practice itself may be less than transparent and not necessarily uniformly consistent.

And even if there exists such familiarity, then many of the inherent limitations of this data source, such as the inability to measure unauthorized migration, can only be explicitly stated as a cautionary note for the readers of the *Outlook* reports. These limitations can unfortunately never be overcome using such data alone.

It is important once more to recall that the preference for residence permit data in the recent OECD *Outlook* reports is motivated by practical empirical considerations, rather than by methodological ones: The two reasons were the desire to have a standardized definition of permanent immigrants that is comparable across OECD member states, and the desire to have an observed set of attributes characterizing the nature of the moves, namely the category under which entry is obtained. Essentially, however, the *counting of persons* who are present in OECD member states and who qualify as migrants under this residence permit definition, remains the approach and the objective of the *Outlook* authors. They make no attempt to focus on the *observation of migratory events* and thereby on the observation of the *migration processes* that are actually taking place.

Yet, at the same time, this new approach in the *Outlook* reports also has at least the potential to be a step in the direction of the modern methods of measurement that form the subject of the present chapter. This is, because when one observes the *timed grant of a residence permit*, one actually observes an *event*, rather than a person.

Put more formally, the OECD authors have the opportunity to improve their methodological perspective by taking the *grant of a residence permit* as a *proxy variable for an -- in itself unobserved -- migratory move*.

Of course, considering the many limitations of residence permit data, it will be obvious that it is preferable to observe such moves in the life histories of persons direct through population censuses or continuous population registration systems. And clearly, the difference between temporary, short-term and long-term migrants is better borne out by observed data on the associated actual residence durations since and between moves, instead of by the nature of any formal residence permits issued.

However, let us pursue the preference of the *Outlook* authors, and let us *assume* that we can adequately heed and accept all the caveats associated with the use of residence permit data.

Then the event-nature of the data can be exploited if the authors are able to establish for each person involved at which point in time in their life history their permit was issued, that is, at which point in time in their life history the event occurred. The simplifying assumption must then be made that the event of granting residency permission and the associated migratory event (the migratory move) may be equated. As we have seen, it is open to questioning whether this assumption may be taken as valid. Further, a second simplifying assumption must be made, namely, that the moves under study do not constitute potentially recurrent events. For the long-term immigrant category selected by the authors, this assumption may reasonably be made, at least over a limited time span of measurement and analysis. Next, for more informative analysis, it is preferred that persons be classified in age-specific (that is, age-homogeneous) cohorts. This is something which the authors currently do not do. Finally, classification by category under which entry is obtained, of course, also remains available as a further attribute by which cohorts may be disaggregated.

Given such sequences of life history data, and given benchmark cohort or population sizes in receiving countries, then the full power of the demographic measurement of migration processes, including, for example, the measurement of their intensities, of the absolute magnitude of the flows and of the resulting stocks, and of their dynamics over time and age, becomes available.

Clearly, one may justifiably criticize the choice of the proxy variable on both conceptual and operational grounds. However, this approach would represent a shift from migrants to migratory moves as the primary object of observation, and a shift away from simple *person-based counting* to a methodologically rigorous and much more informative *event-based analysis*.

In so doing, it would mark a significant departure from the traditional attribute-based counting paradigm discussed in section 4 of this chapter, a paradigm that is typical of current OECD *Trend* / *Outlook* reports. As we have seen in section 4, the simple counting of persons selected on the basis of some attribute which is taken as indicative of past migration behaviour is a paradigm that seriously limits the ability to observe and interpret actual true migration processes. The OECD *Trend* / *Outlook* reports are just one example illustrating that this paradigm has, nevertheless, become deeply entrenched in the applied measurement of international migration.

On the subject of census data processing, finally, we list one UN publication that deserves attention. It dates back to 1999, and, inevitably in the rapidly changing world of information technology, this is evident in the selection and treatment of some issues. However, it contains several observations that are still relevant and valuable today. In addition, these UN guidelines and recommendations are helpful in positioning our discussion in section 6, above, in a broader context against the background of an overall design of the information systems of national and international statistical offices.

United Nations Statistical Commission and Economic Commission for Europe (1999) *Information Systems Architecture for National and International Statistical Offices. Guidelines and Recommendations*. Conference of European Statisticians Statistical Standards and Studies No 51. United Nations

From the perspective of our section 6 it is also interesting to note in particular that, as early as in 1999, the authors already recognize and support the view that relational database management systems, the SQL interface language, and software tools compatible with the relational data model are *de facto* standards for the management and processing of microdata in national and international statistical organizations.

Further information on the *Population & Migration Programme* of ANRC Consulting is available on the internet at:

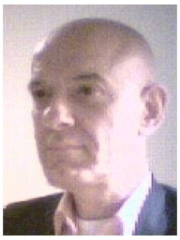
www.anrc-consulting.com/en/csr.html

"**Introduction to the Measurement of Internal and International Migration**" introduces the reader to the state of the art in methods of measuring migration. Against the background of established traditional approaches to measurement, it describes a new generation of methods which, for the first time in the study of migration, are capable of producing *methodologically sound*, *empirically comprehensive* and *fully detailed information* on ongoing processes of internal and international migration. This includes, for example, full information on: migration rates, migrant stocks and flows, and their dynamics over time and age; migration trajectories, including circular and return migration; short-term versus long-term migration; frequent migration; and estimates of and adjustments for the incomplete enumeration of migrants. The book also outlines approaches for in-depth analyses and for explanatory studies of migration behaviour.

In addition to methods of measurement, the book pays detailed attention to the applied context. In the second half it fully covers the key issues, including operational data specifications, sources of data and methods of data collection, and data processing.

As an introduction, the book is readily accessible to readers with a limited or no prior background in migration studies. The focus is on the presentation of practical methods and materials for direct *application*. A prime objective is to facilitate the production of comprehensive, relevant and timely information on migration processes to enable and support evidence-based policy making, planning, monitoring, analysis and evaluation.

This book is an essential practical resource and reference for *national and international statistical offices*, as well as for *research institutes* focusing on population issues. In addition, it is particularly suitable as an introductory textbook for *undergraduate and post-graduate courses* in areas such as demography, population studies, statistics, economics, sociology, geography, and urban and regional planning. And with its applied focus, it is also ideally suited for shorter dedicated *professional training courses and capacity building programmes*.



About the author: William L J Xu-Doeve is a research methodologist, mathematician and demographer with over 30 years of experience in census and survey research, data processing, demographic analysis, modelling and forecasting. In 1982 he was elected to the International Union for the Scientific Study of Population (IUSSP), and he is a leading expert in the field of migration. He has held positions at a range of universities, in areas as varied as physics, econometrics, statistics, demography, geography, development planning, and business studies. He advises governments and the private sector worldwide on strategic management and on information technology. He is actively involved in structural capacity building programmes in developing countries, focusing on data collection and official statistics, and on the production of relevant and timely information, including on internal and international migration.

ISBN 978-90-8802-002-5

9 789088 020025

ISBN: 978-90-8802-002-5
Price: € 87.50